MULTISOURCE INFORMATION COORDINATION:
APPLICATION AND EVOLUTION OF BIG DATA IN CITY AND REGION

多源信息协同

——城市和区域级大数据的应用与演进

贾晓丰　梁郑丽　任锦鸾　著

清华大学出版社

多源信息协同
——城市和区域级大数据的应用与演进

MULTISOURCE INFORMATION COORDINATION IN COMPLEX GIANT SYSTEM:
APPLICATION AND EVOLUTION OF BIG DATA ON CITY AND REGION CLASS

贾晓丰　梁郑丽　任锦鸾　著

清 华 大 学 出 版 社

北　京

内 容 简 介

本书系统地介绍了复杂巨系统下的多源信息协同体系，主要内容分为三篇：上篇为全球趋势篇，介绍智慧城市大数据的战略诉求与协同应用；中篇为协同体系篇，介绍多源信息协同的标准体系、自适应模式与总体架构设计；下篇为内生结构篇，介绍多源信息协同网络的差异测度和关系测度，及其相应的模式优化策略。

本书内容聚焦城市化和信息化的交叉领域，涉及城市管理、信息科学、计算科学、系统科学、决策科学、模糊数学、统计学、社会网络分析等多个学科。本书给出了相关案例的核心程序和关键计算数据，所有代码、数据均经过严格测试。

本书可作为高等院校管理科学、系统工程、信息技术等相关专业研究生和高年级本科生参考用书，也可供相关科研人员、管理人员和工程技术人员参考。

图书在版编目(CIP)数据

多源信息协同——城市和区域级大数据的应用与演进/贾晓丰，梁郑丽，任锦鸾著.
—北京：清华大学出版社，2016
ISBN 978-7-302-42649-3

Ⅰ.①多…　Ⅱ.①贾…　②梁…　③任…　Ⅲ.①信息系统－研究　Ⅳ.①G202

中国版本图书馆 CIP 数据核字(2015)第 320682 号

责任编辑：焦　虹
封面设计：傅瑞学
责任校对：梁　毅
责任印制：刘海龙

出版发行：清华大学出版社
网　　址：http://www.tup.com.cn，http://www.wqbook.com
地　　址：北京清华大学学研大厦 A 座　**邮　　编**：100084
社 总 机：010-62770175　**邮　　购**：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
课件下载：http://www.tup.com.cn，010-62795954
印 刷 者：三河市君旺印务有限公司
装 订 者：三河市新茂装订有限公司
经　　销：全国新华书店
开　　本：170mm×230mm　**印　　张**：19.25　**字　　数**：285 千字
版　　次：2016 年 4 月第 1 版　**印　　次**：2016 年 4 月第1 次印刷
印　　数：1～1500
定　　价：49.00 元

产品编号：067300-01

本书受“国家自然科学基金”(71172040)资助

This monograph was sponsored by

National Natural Science Foundation of China（71172040）

前 言

自人类结绳记事起，数据即如涓涓细流，润物无声地流淌过历史的每一个瞬间，产生抑或湮灭，无止无息。**人类的历史是数据的历史，人类的未来亦是数据的未来**。

近年来，随着新一代信息技术的应用和全球智慧城市的规模化发展，人类社会和虚拟世界的边界迅速向物理空间延伸，各行各业的数据呈现爆发式增长，“人-机-物”三元世界的融合正在加速形成。大数据的“魔盒”一经开启，传统行业的颠覆随之到来。如何应对信息革命带来的社会变革，并借助历史潮流的趋势，推动人类社会在变革中持续进步，是每一位数据科学家、经济学家、系统学家、社会学家、企业家及相关行业的管理者和从业者所共同面对的重大课题。

钱学森先生在1990年自然杂志第1期中发表了著名文章《一个科学新领域——开放的复杂巨系统及其方法论》。智慧城市作为一个“开放的复杂巨系统”，城市本身与周围的环境有物质、能量和信息的交换，同时城市系统下又包含数量庞大、种类繁多的子系统。随着城市管理从“经验治理”向以信息为核心的“科学治理”加速转变，跨部门、跨领域、跨层级、跨主体的多源信息协同是保障城市系统中其他资源要素优化配置的基础，也是城市更加智慧运行的前提。

在科学命题的哲学范畴里，大数据的智慧正如同拉普拉斯宿命论式的畅想：“有一种智能，在任一瞬间里都能识别所有移动的力，以及力与力相互之间的状况。最好是能得到足够巨量的数据来分析，用同一种程序既能分析宇宙中最大的天体的运动，也可以分析最轻的原子的运动。没什么是不能确定的，对于这种分析程序来说，**未来就像过去一样看得清清楚楚**。”

对城市和区域级的复杂巨系统而言，数据集随时处于量级的增长和维度的变化中，原始形态的大数据一般不容易被验证和确认，大数据分析的过程和方法也难以被完整地复制。因此，我们试图从信息协同的视角出发，探讨信息在城市系统中的流转模式，通过信息将人、地、事、物、组织关联起来，形成一个以数据为核心的统一有机体；在此基础上，分析一个复杂巨系统的运行机理和发展规律，并进一步解读城市和区域级大数据的应用和演变之道。

本书内容聚焦城市化和信息化的交叉领域，涉及城市管理、信息科学、计算科学、系统科学、决策科学、模糊数学、统计学、社会网络分析等多个学科。主要分为三篇：

上篇　全球趋势篇(第 1、2 章)：从全球智慧城市规模化发展和新一代信息技术普及应用的视角出发，介绍智慧城市大数据的战略诉求与协同应用。

中篇　协同体系篇(第 3～6 章)：以智慧城市为对象，介绍复杂巨系统下的多源信息协同的标准体系、自适应模式和总体架构设计。

下篇　内生结构篇(第 7、8 章)：介绍多源信息协同网络结构的差异测度和关系测度方法，及其相应的模式优化策略。

本书原则上不对基础理论和技术工具的应用进行普及性介绍。为保证理论体系的完整性，对群体决策、模糊聚类、凝聚子群等关键概念和方法进行必要描述。感兴趣的读者请自行查阅相关专业书籍和研究文献。

本书对外部内容和观点的引用按原始出处列入参考文献，在此向所引内容和观点的作者表示感谢。如有因各种原因造成的引注错误或疏漏，请广大读者及时指出。为了方便理解和开展进一步研究，本书给出了相关案例的核心程序和关键计算数据，欢迎读者合理使用并规范引注。

本书受**“国家自然科学基金”**(**71172040**)、**“中国传媒大学优秀中青年教师培养工程”**项目支持。全书由贾晓丰主持编写和统稿。

本书付梓之际，特别感谢中国科学院的陈锐研究员、赵宇博士、王宁宁

博士及工信部情报所、清华大学、北京大学、北京航空航天大学、中国人民大学、首都经济贸易大学的相关专家和学者提出的宝贵意见。

限于作者的学识水平，书中错误及不妥之处在所难免，恳请广大读者和业界同仁不吝赐教。希望本书能够为智慧城市信息化顶层设计和区域级信息协同的理论与实践起到一定参考作用，为我国城市化和信息化事业的蓬勃发展尽微薄之力。



2016 年于中国科学院



目 录

上篇　全球趋势：
智慧城市大数据的战略诉求与协同应用

下篇 内生结构：多源信息协同网络的测度与优化

上篇　全球趋势：

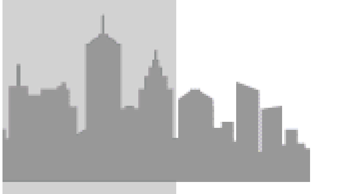
智慧城市大数据的战略诉求与协同应用

信息化水平是当代社会生产力的重要标志，信息化的终极目标是利用信息技术将人类社会与信息空间、物理世界相互融合，形成“**人-机-物**”**三元一体**的世界。作为城市发展的高级阶段，智慧城市正在从“经验治理”向以信息为核心的“科学治理”加速转变，城市大数据的协同应用是转变过程中解决“**智慧孤岛**”、重构产业格局的核心助推器。

第 1 章　智慧城市与城市大数据

1.1　全球智慧城市规模化发展

随着信息网络的高速发展，资本和劳动力的全球流动性增加，大规模的城市化运动在全球展开。根据 2010 年 3 月 25 日联合国经社事务部人口司在纽约总部发布的《世界城市化展望 2009 年修正版》报告，到 2050 年人口总数将达到 97 亿人，生活在城市中的人口将达到 64 亿人，中国的城市人口预测超过 70％。全球城市化在推动经济社会发展的同时也带来了许多无可逃避的问题，如城市人口拥挤、工业污染、垃圾污染、交通拥堵、资源短缺等种种城市病已经成为影响城市未来发展的重要障碍，其根本原因在于传统城市在社会管理和服务上的滞后。为解决这些问题，实现社会经济的可持续发展，全球许多国家提出了智慧城市建设，城镇化和信息化成为当今时代的发展主题。智慧城市是人类文明的集中体现和综合应用，是当代城市发展的崭新模式，也是城市信息化发展的高级阶段，已在全球范围内成为一种趋势。



1.1.1　关于智慧城市内涵的讨论

目前，业界对于智慧城市内涵的探讨大致分为三类：

第一类侧重技术的重要性，认为智慧城市是信息化、工业化和城镇化高度融合的结果，智慧城市中信息技术呈现泛在化、效用化、智能化、绿色化和软性化的“五化”特征，强调通过新一代信息技术来实现城市感知、互联和智能的特性，使整个城市系统高效、智能、协调运作。

第二类侧重最终实现的愿景，认为智慧城市的概念有狭义和广义之分，技术只是智慧城市的一种实现手段，智慧城市的完整内涵应该涵盖全民参

与、城市居民生活质量的提高、个性化需求的满足、城市生态空间的开放与创新等多个方面。

第三类侧重“城市化”与“信息化”的结合，认为智慧城市应该建立在一种“具有思维的”且“内生性的”科技与社会相互作用的过程之中，将智慧城市同技术发展与资源观念的演变联系在一起。

在应用层面上，目前国内外对智慧城市的理解主要聚焦在三个层面：一是空间维度，重点体现在智慧社区和智慧园区的建设；二是行业维度，主要体现在教育、卫生、文化、旅游、航运、物流等不同领域对未来发展的新需求；三是管理维度，主要体现在如何为整体的智慧城市建设营造更好的环境。

综合来看，智慧城市就是通过新一代信息技术，迅速、灵活地处理各种事项，实现城市智慧化运行与管理的状态。在本质上，智慧城市是城市经济发展模式创新在特定空间上的具体体现，是城市发展的新模式和新形态，是人类在城市化进程中，实现人与人、人与城市和自然高度融合、协调发展的更高阶段，通过全信息链应用体系，使城市具有正确理解和处理政府、企业、居民所涉及的各种活动的能力，实现互联互通、迅速灵活、高效优质、协同共享的目标。

1.1.2 全球视野下的差异化发展与共性聚焦

自从 2009 年初 IBM 提出“智慧地球”的概念以来，美国、日本、荷兰、英国、瑞典、韩国、新加坡等发达国家和地区相继发布了建设智慧城市的相关规划和政策，全面开展智慧城市的建设。

（1）美国：美国提出了加强智慧型基础设施建设和推进智慧应用项目的经济刺激计划，借助于发展智能电网等基础设施，致力于培育更多的新兴产业和新兴服务。其中，纽约市在应急保障与社会安全体系建设、圣地亚哥市在智能电表和清洁能源应用、迪比克市在水、油、交通等城市资源协同服务等方面均取得了显著成效。

（2）欧盟：欧盟制定了智慧城市框架，重点加强在气候问题应对和节能



减排方面的投入，提出“20/20/20 by 2020”的战略目标，即相对于1990年，在2020年实现温室气体减排20%，将可再生能源的使用比率提高20%，节能20%。为实现三个20%的目标，欧盟各大城市纷纷加快启动智慧城市建设示范项目。

(3) 亚洲：亚洲以韩国、日本和新加坡为代表，在大力开展国内智慧城市建设的基础上，面向新兴经济体出口智慧城市产品和解决方案。韩国、日本先后在U-Korea、U-Japan的国家战略规划基础上推出了进一步的升级计划。韩国以泛在网络为基础，在首尔、松岛等地开展了U-City智慧城市建设试点；釜山的云计算即付即用模式、济州岛的智慧岛项目等均是韩国基础设施与服务出口的典范。日本的智慧城市建设涵盖新能源汽车、智能电网、智能家庭、节能环保等多领域的“多元化发展”，强调跨领域的协同合作。例如横滨在能源、建筑、交通等领域协同发展，通过引入新技术降低碳排放。除此之外，新加坡提出2015年建成“智慧国”的计划，台湾地区提出建设“智慧台湾”的发展战略等，均是在结合地区特色和战略定位的基础上，寻求各自智慧城市发展的切入点。



当前，全球的智慧城市建设在规模化扩张的同时，已逐渐凸显出各自的发展方向和区域特色，如维也纳的智能电网、多伦多的循环经济、东京的移动智能、伦敦和瑞典的智慧交通、巴黎的自行车共享、德国的电动汽车、哥本哈根的创新清洁技术、巴塞罗那的光伏产业等。然而从全球视野来看，不同国家和地区的差异化发展过程中，又进一步呈现出了领域重点的聚焦，如社会服务的智能化与个性化以及城市治理的协同开放。

1.1.3 国内智慧城市建设的政策背景与信息壁垒

根据工信部2013年第1号通告，截至2012年年底我国已有320个城市投入3000亿元建设智慧城市，智慧城管、智慧e通等一系列信息化服务走进百姓生活。北京、上海、广州、深圳、宁波、扬州、杭州、南京、海口等各大城市均结合当地区域的特点和需求，制订了各自的发展规划；同时，在城市普遍

面临的各类“城市病”和关键问题上，也显现出了一定的共识。

2013年8月，国务院发布的《关于促进信息消费扩大内需的若干意见》，明确提出“加快智慧城市建设”。同年，工业和信息化部等八部委联合起草了《关于促进我国智慧城市健康发展的指导意见》的征求意见稿。2013年1月29日，住房和城乡建设部公布了首批国家智慧城市试点名单，共90个城市；8月5日，公布了第二批名单，共103个城市（区、县、镇）试点；2015年4月7日公布了第三批名单，共84个城市（区、县、镇）试点及13个城市（区、县）扩大范围试点。2013年11月21日，中欧城镇化伙伴关系论坛分别确定了中欧15个试点城市，共同作为中欧智慧城市合作试点城市。2014年8月，国家发展改革委等八部委联合下发《关于促进智慧城市健康发展的指导意见》，进一步推动和规范智慧城市建设进程。智慧城市已成为拉动城市升级、经济转型和改善民生的战略选择。

国务院副总理马凯同志在2014年2月18日召开的全国物联网工作电视电话会议中明确要求“扎实推进物联网有序健康发展，在食品安全、社会保障、医疗卫生、城市管理、民生服务、公共安全等领域开展应用示范”。以物联网、云计算、移动互联网、大数据为代表的新一代信息技术对推动创新浪潮和产业革命、建设现代信息技术产业体系具有重大战略意义。

当前，我国的智慧城市建设处于基础设施建设和领域示范应用的起步阶段，涉及社会管理、应用服务、基础设施、智慧产业、安全保障、建设模式、标准体系等内容，智慧城市的架构模式、标准规范、关键技术、评价体系等均不成熟。作为后IP时代跨越信息壁垒的关键突破口，以物联网为引领的智慧应用建设在解决一个个信息孤岛的同时，不可避免地又形成了领域间的新的智慧孤岛。信息协同对于智慧城市大数据管理的重要性日益凸显。进入大数据时代，智慧城市建设的关键不再是数字城市建设中的信息化系统，而是面向城市和区域系统下的多源信息的实时融合，在城市范围内实现跨领域的信息协同共享，支撑跨部门的协同联动和智慧城市的精细化管理。

1.1.4 智慧城市评价体系百家争鸣

目前，业界关于智慧城市的评价指标体系尚未形成统一的标准和共识，普遍缺少有效的定量依据和经过实证检验的模型支撑。

国外方面，欧盟从智慧产业、智慧民众、智慧治理、智慧移动、智慧环境和智慧生活 等六个维度对智慧城市建设进行了评价研究；智慧社区论坛(ICF)从宽带连接、知识型劳动力、创新、数字包容、营销和宣传等五个维度对智慧社区的发展水平进行了定性的评估；Boyd Cohen 认为智慧城市是借助信息通信技术来发挥其重要作用的，为城市创新和环保经济提供了支撑。智慧城市可以通过降低城市运行成本、节约资源、减少环境污染来提高城市服务水平和居民生活质量，提出城市创新与城市化可持续发展将成为智慧城市的评价标准。



国内方面，上海浦东于 2011 年 7 月发布了首个中国版本的智慧城市指标体系，包含 19 个二级指标和 64 个三级指标，涉及城市基础设施、公共管理和服务、城市信息服务经济发展、人文科学素养、城市民主感知等五个维度；此外，陈铭、李贤毅、李健、顾德道等学者从不同的角度提出了智慧城市发展水平评价指标体系，其中一级指标主要集中在智慧基础设施、智慧应用、智慧产业、智慧人群、智慧服务等方面。

1.2 城市大数据与新一代信息技术应用

1.2.1 全球加速迈入大数据时代

随着科学技术的进步和人类社会信息化进程不断推进，数据产生成本的下降、投资规模的增加和数据存储能力的增长，使人类所面临的数据量呈现出前所未有的爆炸性增长。

1. 无所不在的数据增长源

不知不觉中,数据增长源已遍布我们每个人的周围:

- 社交网络、电子商务网站、视频网站等互联网应用和服务产生了大量数据。2011 年被创建和被复制的数据总量为 1.8ZB(1ZB=1024EB),远远超过人类有史以来所有印刷材料的数据总量(200PB)。例如 Facebook 每月上传的照片超过 10 亿张,每天生成 300TB 以上的日志数据;淘宝网会员超过 3.7 亿人,每天交易数千万笔,产生约 20TB 数据。
- 物联网和移动计算蓬勃发展产生规模更加巨大的数据洪流。预计至 2020 年,全球将有 500 亿个终端感知设备连入互联网,产生的流数据量将十分惊人。
- 科学研究(如基因组学、天体物理学和脑科学等)也产生了大量数据。例如,用电子显微镜重建大脑中的突触网络,1 立方毫米大脑的图像数据就超过了 1PB。

IDC 认为,全球数据增速符合大数据摩尔定律(又称新摩尔定律),大约每两年翻一番。预计到 2020 年,全球数据量将达到 35ZB,年均增长率则超过了 40%,是 2010 年的 29 倍,如图 1.1 所示。

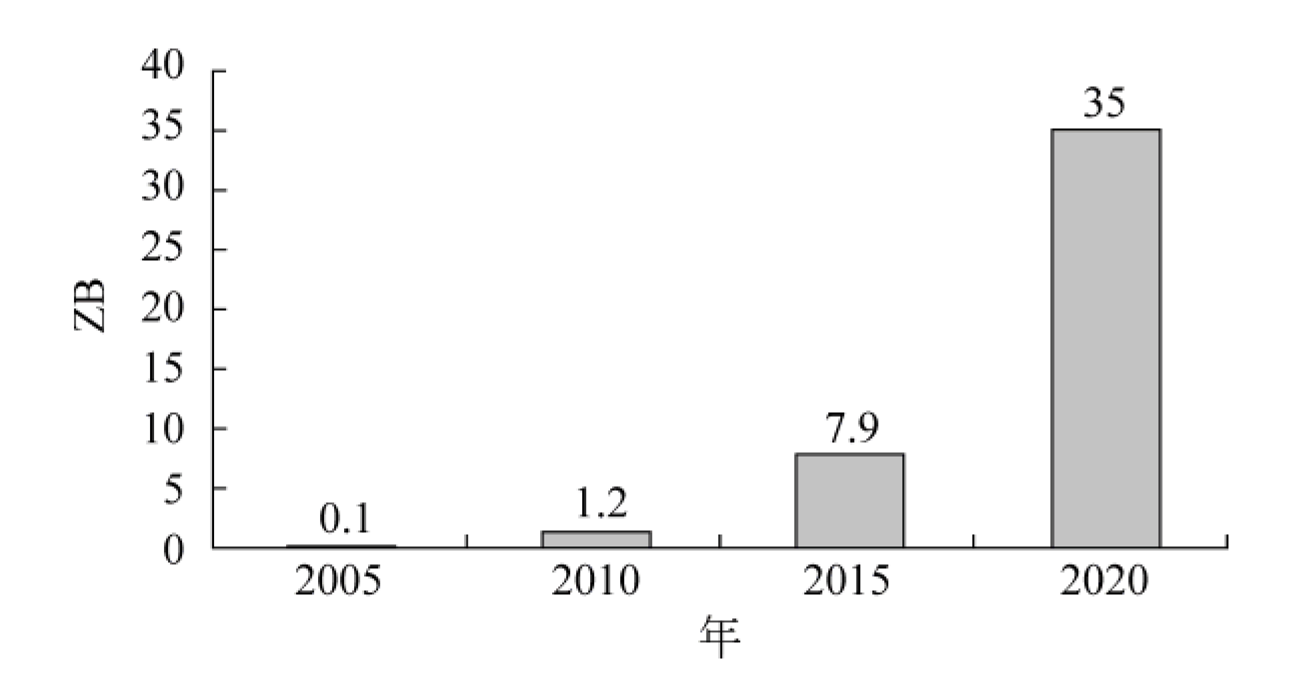


图 1.1　2005—2020 年全球数据量及预测

庞大的数据量及其处理和应用需求催生了"大数据概念",也预示着大数据时代的来临。按维基百科的定义,大数据是指规模巨大到无法通过目



前主流软件工具在合理时间内实现获取、管理、分析挖掘的数据。大数据概念是数据对象、技术与应用三者的统一，其基本特征包括四个方面（即通常所说的4V）：数据规模巨大（Volume），数据类型繁多（Variety），数据处理速度要求快（Velocity）、数据价值密度低（Value）。

2. 全球数据的分布与增长

根据2012年12月IDC发布的数据，美国拥有全球最多的数据（32%），其次分别为西欧（19%）、中国（13%）和印度（4%）。全球其他国家和地区拥有剩下的32%，与美国一国所拥有的数据量大致相当。全球数据地理位置分布如图1.2所示。

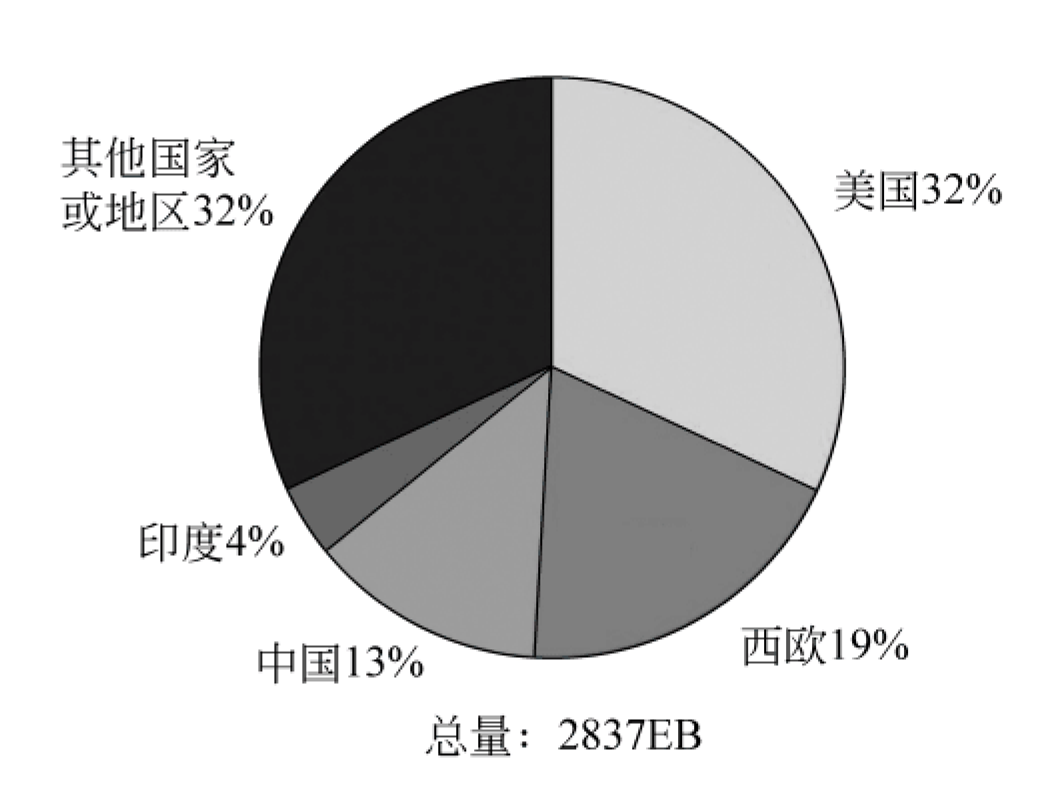


图1.2 全球数据地理位置分布

（资料来源：IDC's Digital Universe Study，sponsored by EMC）

基于互联网资源和技术优势，美国已具备对全球网络空间的监视控制能力。斯诺登披露的“棱镜”计划，缘于美国政府的“星风”监视计划。2004年，布什政府通过司法程序，将“星风”监视计划分拆成由国家安全局执行的4个监视计划，除“棱镜”外，还包括“主干道”、“码头”和“核子”。其中，“棱镜”用于监视互联网个人信息；“核子”则主要负责截获电话通话者对话内容及关键词；“主干道”和“码头”分别对通信和互联网上数以亿兆计的“元数据”进行存储和分析。“元数据”主要指通话或通信的时间、地点，使用的设备及参与者等，不包括电话或邮件等的内容。

根据麦肯锡全球研究中心 2010 年的数据，全球新增数据量地理分布如图 1.3 所示。

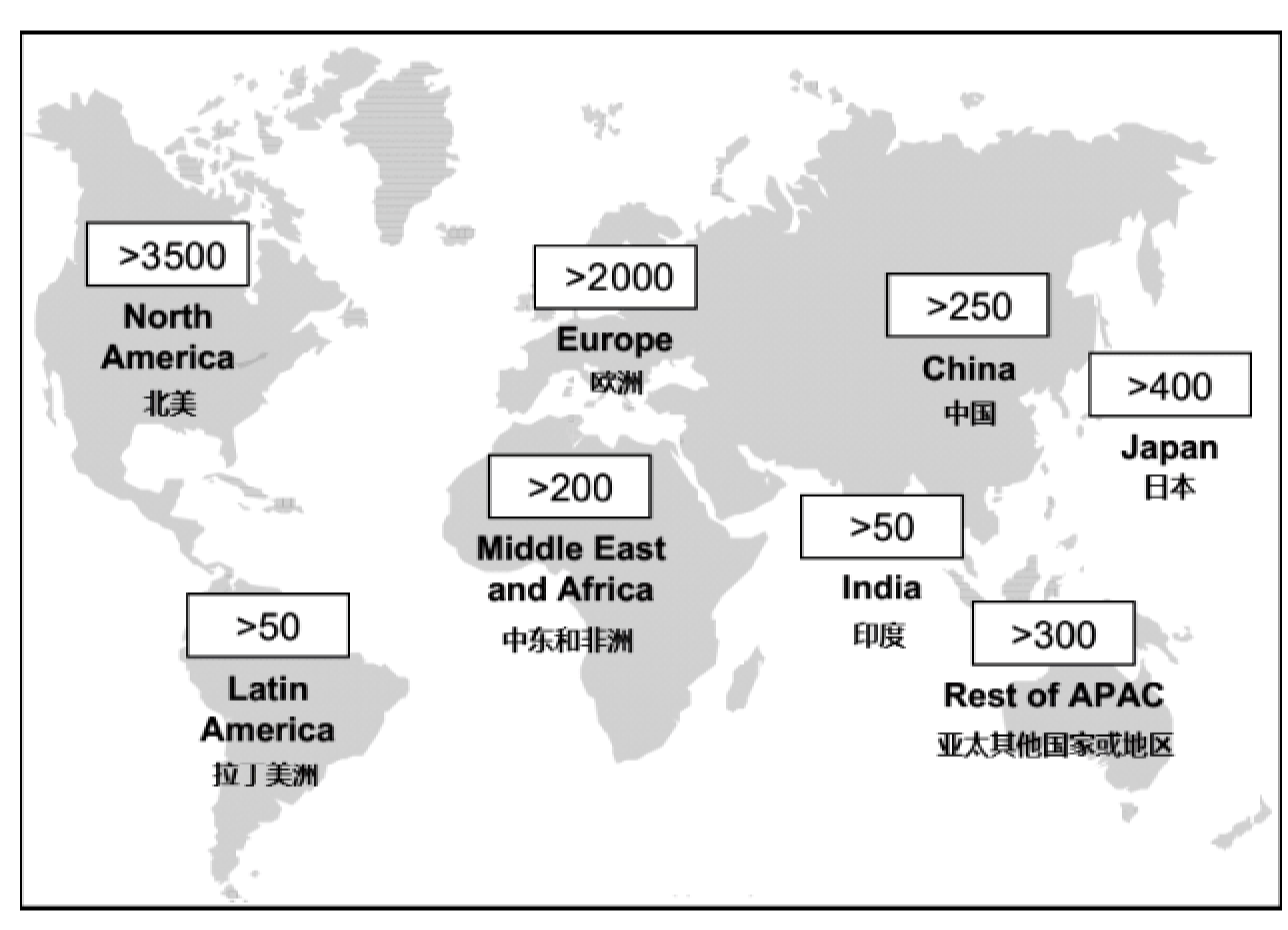


图 1.3　全球新增数据量地理分布

（资料来源：IDC storage reports；McKinsey Global Institute analysis）

我国拥有全球第一的人口数、互联网用户数和移动互联网用户数，数据存量和潜在增量位居世界前列。互联网和各行业信息化系统经过多年发展沉淀下来的数据量已经十分庞大。2012 年，中国的数据量为 364EB，占全球 13%，仅电信、金融、医疗、文化等国家重要基础数据总量就超过 900PB。2013 年 1～9 月，我国移动互联网接入流量累计完成 927PB，同比增长 67.3%；2012 年淘宝网每天交易数千万笔，单日数据产生量超过 50TB；百度存储网页数量已接近 1 万亿页，每天处理 60 亿次搜索请求，每日新增数据 10TB，处理数据超过 100PB；上海证券交易所每秒处理近 9 万笔业务，每日成交 3 亿笔以上；中国联通用户上网记录每月 1 万亿条，产生数据 300TB。

未来，随着制造业升级改造不断推进，金融、交通、电信等重点行业和医保、社保、海关等重要领域的业务数据不断集中，我国数据存量将持续快速增长。预计到 2020 年，仅中国就将产生全球 21% 的数据，数据量超过 8ZB，年均增长接近 50%。



1.2.2 城市计算与城市大数据

以物联网、云计算等新一代技术为核心的智慧城市建设理念，已成为一种未来城市发展的全新模式，也是当前全球城市发展的热点。智慧城市是人类社会发展的必然产物，智慧城市的建设有利于解决城市发展问题，提升城市信息管理水平，促进国家高端产业发展。城市计算是智慧城市背景下，城市化与信息化结合的一个新兴领域。在对城市计算的概念进行界定之前，首先对与城市计算相关的新一代信息技术进行定义。

（1）物联网：指通过射频识别（Radio Frequency Identification，RFID）、红外感应、卫星定位、激光扫描、视频监控等信息传感设备，按照约定的协议，把物体与互联网连接起来，进行信息交换和通信，以实现智能化识别、定位、跟踪、监控和管理的一种网络或智慧管理环境。物联网是在互联网基础上的延伸和扩展，具有技术融合度高、产业链条长、应用领域广等特点，一般包括信息的采集（“感”）、传输（“传”）、分析（“知”）、应用（“用”）等多个环节。



（2）云计算：是一种基于互联网的软硬件服务模式，旨在通过最小的管理代价和可配置的计算资源（如服务器、存储器、网络、应用、服务等）为用户提供快速、动态易扩展的虚拟化资源服务。用户只需有简易的终端设备，即可使用浏览器进行身份验证后应用软硬件服务（桌面系统、开放平台、应用系统等），软硬件及数据都在云计算中心。云计算的效率和低成本服务传递优势从技术实现层面为智慧城市的数据中心建设提供了良好的解决方案。

（3）关联数据技术是一个语义网技术的最佳实践，它采用资源描述框架（Resource Description Framework，RDF）数据模型，采用统一资源标识符（Uniform Resource Identifier，URI）命名并生成实例数据和类数据，在网络上进行发布和部署后能通过超文本传送协议（Hypertext Transfer Protocol，HTTP）获取，构建数据互联与人机理解的语义环境。

城市计算的内涵在于将城市空间中的每个传感器、设备、人、交通工具、建筑物、道路等都当作一个单元去感知城市动态，协同完成一个城市级别的

计算以服务于市民和城市。城市数据是多样性和动态性的。例如以城市中的出租车为单元，可以基于出租车速度的分析挖掘道路上的交通热点，基于出租车 GPS 轨迹数据分析，进行两地间的通行时间与费用预测、最优路径选择和道路规划问题评估等；结合城市中的市民在医疗、社交等各个方面的行为数据，通过城市计算提供流行病预警与趋势分析、旅游推荐和广告投放等服务。

城市计算旨在通过城市感知、数据挖掘、智能提取、改善服务四个环节形成的循环过程来智慧型地提升市民生活和城市环境，以及通过整合交通流量、人口流动、地理和地图数据、环境、能源消耗、人口总数和经济状况等一系列异构数据源来深度分析突发现象背后的本质和科学规律。

大数据与智慧城市有着密不可分的联系。作为城市发展的高级阶段，智慧城市创造了以互联互通、整合共享、协同联动、创新发展为主要特征的城市发展新模式，大数据成为全球信息化的重点所在。智慧城市建设带来了数据量的爆发式增长，城市中密布的各类传感器、移动电话、GPS 设备，甚至城市中的人都可成为信息的源头。目前，智慧城市建设所产生的数据量已超过了 200PB，而大数据就像血液一样遍布智慧交通、智慧医疗、智慧生活等智慧城市建设的各个方面，城市管理正在从“经验治理”转向“科学治理”。智慧城市是否真正“智慧”源自城市大数据，如何挖掘海量数据的潜在价值并为城市系统的运行管理决策提供支撑，是智慧城市建设的关键。

城市基础运行的智能化程度是体现一个城市是否“智慧”的重要指标。本书内容中将多次提及城市基础运行领域，并以此为例进行解构分析。城市基础运行主要指城市基础设施（如道路、桥梁、矿体、水库、地下管网等）及其相关城市部件（如车辆）的运行。城市基础运行管理的主体是政府部门，服务对象面向城市系统中的政府、企业（社会团体）和个人。智慧城市基础运行管理需要对城市部件的基础状态和运行状态进行实时感知和控制，整合城市基础运行相关的政务信息资源和社会信息资源（包括社会公开信息资源及政府有权限提取的非公开信息资源），实现跨领域、跨部门、跨层级、

跨主体的信息共享和业务协同，并借助数据挖掘、系统仿真、智能检索等技术手段，为城市基础运行管理和决策提供有效支撑。

1.2.3 新一代信息技术助力三元世界融合

1. 物联网与移动终端催生城市大数据需求

大数据时代最大的特征不在于数据本身，而是在数据的源头。数据的主要来源不再是普通的 PC 和服务器，而是被物理世界不断创造出来，并被物理世界和生活在物理世界中的人所接收、处理和利用。物联网产生的是物理世界的感知数据，移动终端产生的是人类社会的应用类数据和行为类数据。随着物联网与移动终端的普及和发展，人类社会与物理世界日益紧密相连，大数据在这个过程中应运而生。

2. 云端的选择为大数据决策指引新的航向

云计算和云存储的应用使能够“理解数据、做出决策”的大数据技术成为现实。通过把数据存储和数据分析变成可以更加方便获得的网络服务，全球政府、企业和个人使用、消费信息技术的模式正在改写。借助“云”的伸缩性，构建云端之上的大数据平台，实现大数据资源的“按需配置”，并最终获得更大空间的决策弹性。但是，云端的大数据应用目前仍然存在障碍，如美国能源部提出的数据分析问题，一个基于云端的解决方案无法满足对 EB 量级的数据处理需求。

3. 三元世界下的新一代信息技术趋向融合

在复杂性科学视野下，科技创新必须实现技术发展与应用创新的并驾齐驱。中国科学院战略性科技先导专项“面向感知中国的新一代信息技术研究”中指出，信息化水平是当代社会生产力的重要标志，信息化的终极目标是利用信息技术将人类社会与信息空间、物理世界相互融合，形成“人机物”三元一体的世界。大数据与物联网、云计算、移动互联网等新一代信息技术共同构成“人机物”三元世界融合的助推器(参见图 1.4)。

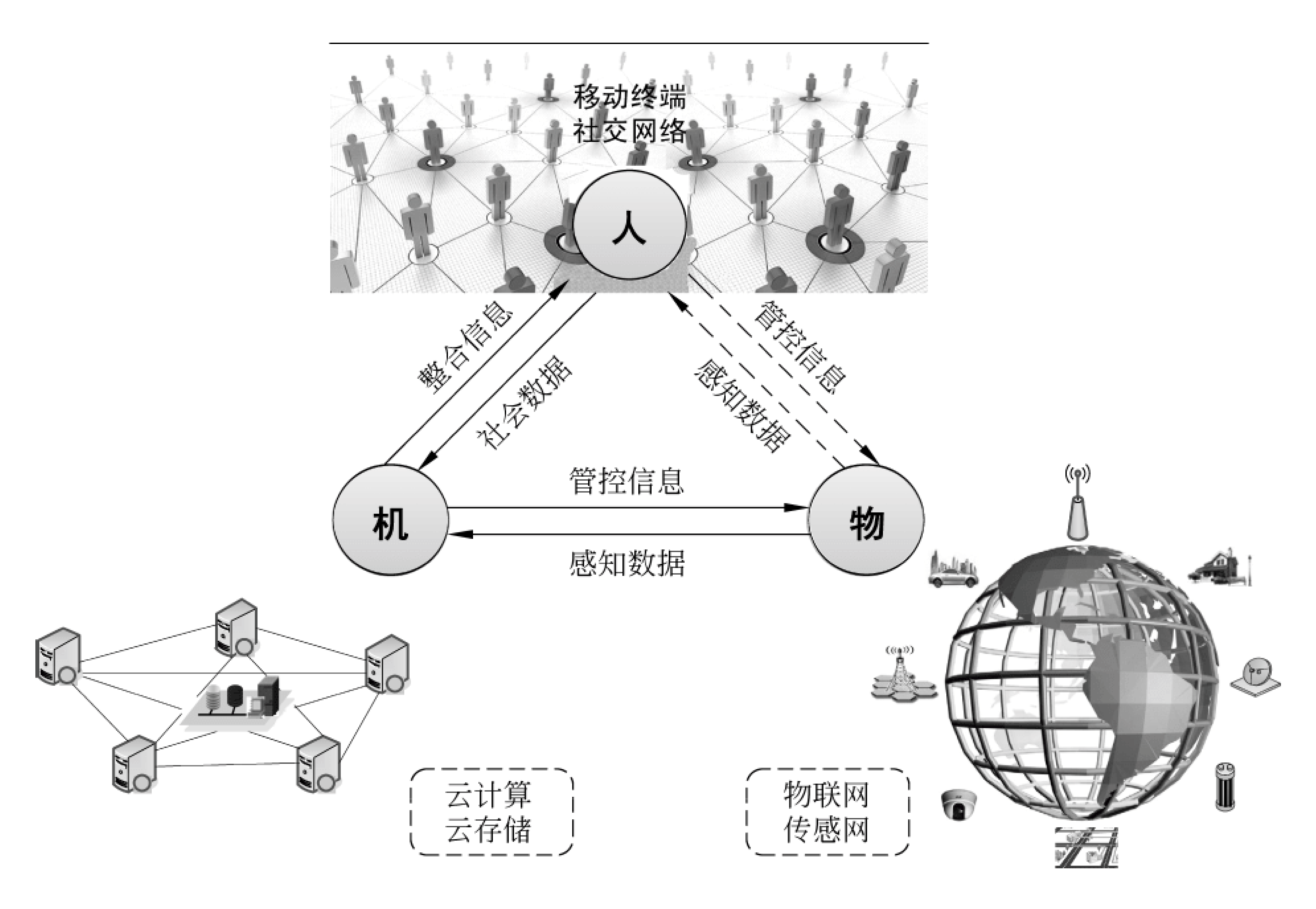


图 1.4 “人机物”三元体系下的新一代信息技术

移动终端颠覆了传统的社会行为和关系网络，物联网融合了人类社会与物理社会的边界，云彻底改变了信息服务的方式，而大数据则重构了相关产业和领域的格局。“人机物”三元融合体系的形成是城市化发展的大势所趋，新一代信息技术的融合则是这次变革浪潮的核心推动力。

1.3 大数据决策：大数据时代的大变革

人、机、物三元世界的高度融合引发了数据规模的爆炸式增长和数据模式的高度复杂化，特别是 Web 2.0、传感器、视频监控等的应用，使得数据量以前所未有的态势迅猛增长，世界已进入网络化的大数据(Big Data)时代。大数据带给世界一个全新的解决问题的方法，将成为引领社会变革、推动政府职能转型、激发企业技术创新的利器。在这个技术浪潮到来之际，如何应时而变是每一个决策主体(组织或者个人)所必须认真面对的问题。

1.3.1 大数据时代的战略诉求

1. 大数据从商业行为上升到国家战略

大数据的应用最初在互联网技术和商业模式发展中崭露头角，商业活动的每一个环节都建立在数据收集、分析和行动的能力之上。IDC 和麦肯锡的大数据研究显示，大数据主要在四个方面带来了巨大的商业价值：一是对顾客群体的细分；二是运用大数据模拟实境，发掘新的需求和提高投入回报率；三是提高大数据成果在各部门间的分享程度，提高企业的整体管理链条和产业链条的投入回报率；四是进行商业模式、产品和服务的创新。

2010 年，英国卡梅伦政府提出“数据权”(Right to Data)的概念，将其作为信息时代向全社会普及的公民基本权利之一。美国奥巴马政府提出“开放战略”，将数据开放作为政府、经济和社会开放的根本动力。这些概念和行动纲领的提出，标志着数据的定位正在逐渐从商业领域走进各国政府的战略核心。



2012 年 3 月，美国政府投资 2 亿美元启动“大数据研究和发展计划”，通过提高从大型复杂的数据集中提取知识和观点的能力，加快在科学与工程中前进的步伐，加强国家安全，推进科学发现和创新研究。这个计划的发布正式将大数据技术从商业行为上升到了国家战略。这是继 1993 年美国宣布“信息高速公路”计划后的又一次重大科技发展部署。美国政府将大数据比作“未来的新石油”，将“大数据研究”首次上升为国家意志。

联合国 2012 年在纽约总部发布了一份大数据政务白皮书，指出大数据时代已经到来，希望各国使用极大丰富的数据资源来更好地服务和保护人民。

2013 年，英国商业、创新和技能部宣布，将注资 1.89 亿英镑研发大数据技术，将在基础设施方面投入巨资，加强数据采集和分析，从而让英国在“数据革命”中占得先机。2013 年的八国峰会(G8 Summit)，签署了《八国集团

开放数据宪章》(G8 Open Data Charter),明确了开放数据的5大原则和14个重点开放领域,其主要宗旨是推动政府更好地向公众开放数据,挖掘政府拥有的公共数据的经济潜力,促进经济增长,激发创新,并加强责任感。

2015年9月,我国国务院正式发布《国务院关于印发促进大数据发展行动纲要的通知》(国发〔2015〕50号),成为我国促进大数据发展的第一份权威性、系统性的文件。该文件从国家大数据发展战略全局的高度,提出了我国大数据发展的顶层设计,是指导我国未来大数据发展的纲领性文件,其核心是推动各部门、各地区、各行业、各领域的数据资源共享开放。中国和美国政府的大数据战略比较详见本书附录A。

大数据带来了深刻的社会变革,能够显著提升政府效率、透明度和服务水平。美国国家安全局(NSA)利用大数据分析来对抗恐怖主义活动,军方利用大数据搜查到拉登的蛛丝马迹,从而将其击毙。政府和社会数据的深度利用,有利于透明化与民主监督,增强公民参与意识,推动自我赋权(self-empowerment),改进政府服务效率和水平,加强政策影响力评估,推动产生新知识,改进或创新产品和服务等。

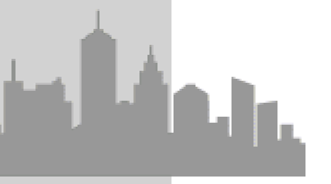
大数据驱动下的信息产业日渐成为关系国民经济和社会发展全局的战略性、先导性产业。拥有数据的规模和质量以及对数据的控制和运用能力,将直接决定一个国家的核心竞争力。大数据像土地、石油和资本一样成为经济、社会运行中的根本性资源,国家的数据主权将是继海、陆、空、天、网之后另一个大国博弈的空间。

面对各国政府纷纷出台政策推动大数据发展,我国也积极应对,在多个科技项目中进行了重点支持。如2011年"核高基"科技重大专项将非结构化数据管理作为需要突破的关键技术加以重点支持;2012年12月国家发展改革委员会将"数据分析软件开发和服务"列入专项指南;2013年大数据被科技部列入"973基础研究计划"和国家自然科学基金指南中;2014年,科技部启动了"大数据计算"重点基础研究发展计划,国家自然科学基金委员会也启动了大数据重点项目群。

2. 大数据从数据资产转变为战略资源

相比传统的海量数据，大数据从 TB 级别跃升到 PB 甚至 EB 级别，在量级上有了更大的提升。除了体量巨大之外，大数据还具有三个海量数据所不具备的特点：第一，数据类型多，音频、视频、图片、GPS 等各类数据广泛汇集；第二，价值密度低，如连续不间断的视频监控过程中，可能仅有几秒的数据是有价值的；第三，处理速度快，这与传统的数据挖掘技术有着本质的不同。

人类从工业时代进入信息时代的一个显著特征，即数据成为政府、企业和个人的重要无形资产，与固定资产共同成为生产过程中的基本要素。随着大数据时代的到来，由于数据量级的指数型增长及其本身蕴含的巨大挖掘价值，大数据的定位已不仅仅是传统意义上的数据资产，而是转变为与自然资源、人力资源同等重要的新型战略资源，辐射到政治、军事、社会、科技、商业、环境等各个领域。



大数据带来思维方式、商业运作和管理制度等多方位的变革，涉及政府、行业企业和个人，是现在和未来的战略制高点。人类第一次有机会和条件，在如此众多的领域和如此深入的层次获得和使用全面数据、完整数据和系统数据，深入探索现实世界的规律，获取过去不可能获取的知识，得到过去无法企及的商机。通晓如何利用大数据的国家或企业将具备新的竞争优势，重新划定竞争版图。

3. 大数据从智能分析延伸到科学决策

随着新一代信息技术的兴起，物联网、移动终端、社交网络、GIS 等的广泛应用为大数据提供了丰富的数据来源。数据中包含着每个用户的身份、地点、时间、喜好、厌恶、行为、社会关系等大量直接或潜在的信息。随着数据挖掘技术的发展，面向大数据的智能化分析不可避免地成为了科技界和企业界共同关注的前沿热点。

在思维方式上，数据的丰富及易得将改变人类认识世界的方式。

（1）从样本式推导走向全数据审视。过去的科学家、社会学家、经济学家、企业家等由于技术和资源的限制，只能通过采样调研和统计分析等手段了解关注的对象，而如今数据的采集和存储成本已经很低，完全可以通过全部数据进行分析，不存在样本抽样的概念。大数据已成为继实验归纳、模型推演和计算机模拟等范式之后的第四科研范式①。

（2）从精确性走向混杂性。大量数据的应用还具有充分的容错性，过去采样的数据如果出现失误可能导致统计结果偏离严重，而采用全部的数据则会将有瑕疵的若干数据淡化处理。

（3）相关关系成为因果关系的有效补充。数据分析不再局限于验证已有的推测是否正确，而是努力寻找背后的因果关系。在很多情况下，只需要知道“是什么”就能做出决策，而不需要对“为什么”投入大量人力、物力进行探究。

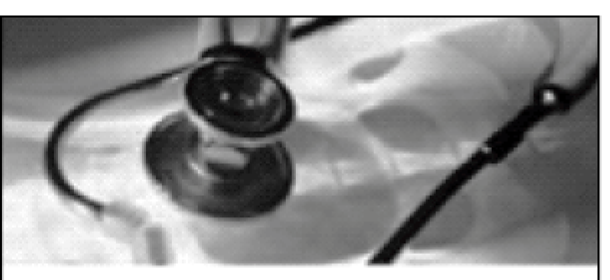
大数据将产生新知识，促进创新，推动传统产业转型发展，催生全新产业，产生巨大的经济价值，成为产业升级与经济转型的创新要素。数据的重新组合将会创造新的知识和思想，甚至创造全新的领域。比如在 19 世纪，研究人员通过将黑死病死亡率和饮用水井的地理分布联系起来，发现了饮用水污染和黑死病之间的关系，从而推动伦敦建造了全新的排污系统，大幅度改善了公众卫生状况。

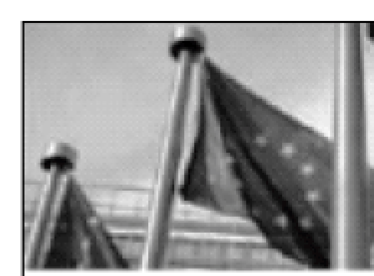


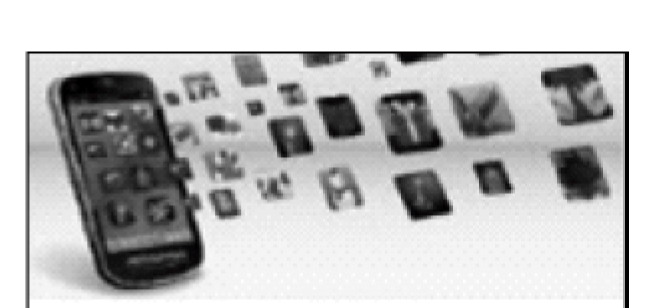
通过大数据的重新组合和深入应用，人们可以期待发现更多类似的新知识。据麦肯锡统计（见图 1.5），大数据能为美国医疗服务业每年带来大约 3000 亿美元的商业价值；能为欧洲的公共管理每年带来 2500 亿欧元的价值，能帮助美国零售业提升 60％的净利润，并帮助降低美国制造业 50％的产品开发、组装成本。美国通用电气公司通过每秒分析上万个数据点，融合能

① 范式（paradigm）的概念和理论由美国著名科学哲学家托马斯·库恩（Thomas Kuhn）提出并在《科学革命的结构》（*The Structure of Scientific Revolutions*）（1962）中系统阐述，指的是一个共同体成员所共享的信仰、价值、技术等的集合。它是常规科学所赖以运作的理论基础和实践规范，是从事某一科学的研究者群体所共同遵从的世界观和行为方式。

量储存和先进的预测算法，开发新型风机，效率与电力输出分别比现行风机提高了25％和15％。Gartner预测，大数据将为全球带来440万个IT岗位，1300多万个非IT岗位。数据使用率提升10％对行业人均产出的平均提升幅度如图1.6所示。











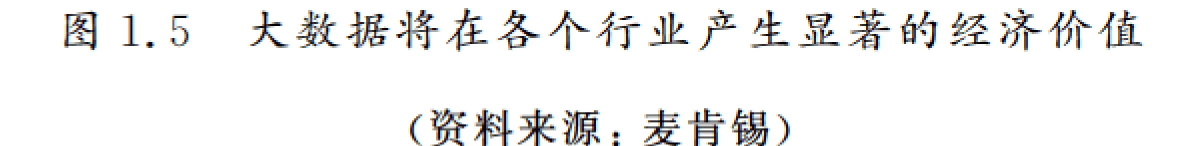


图1.5　大数据将在各个行业产生显著的经济价值

（资料来源：麦肯锡）

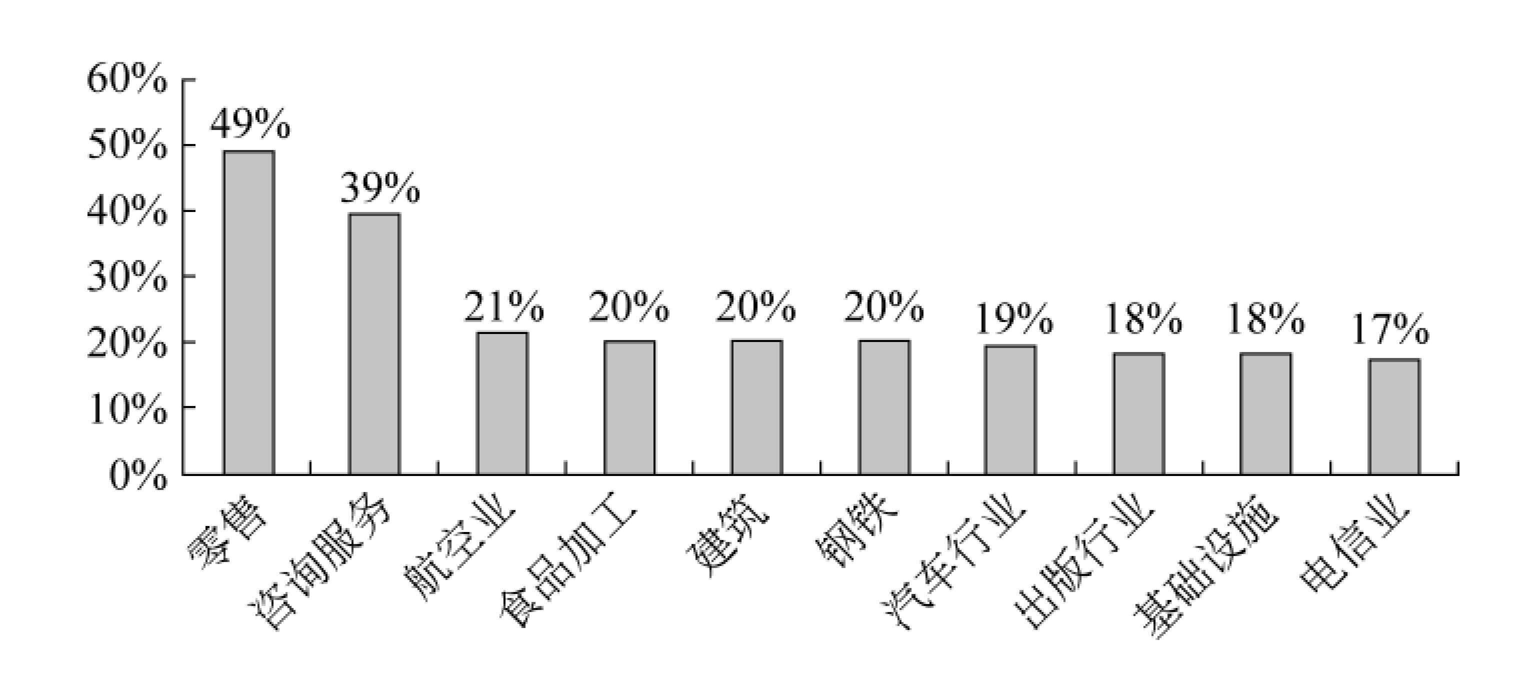


图1.6　数据使用率提升10％对行业人均产出的平均提升幅度

（资料来源：美国得克萨斯大学 *Measuring the Business Impacts of Effective Data*）

大数据深刻影响着每一个人，更好地使用大数据可以帮助个人降低信息获取成本，在生活中做出更好的决策，增加社会活跃度，提升生活质量。

如美国政府数据门户网站(Data.gov)提供50多类数据以及处理这些数据所需的软件工具,所有人都可以自由下载使用。该网站的数据资料不仅有利于公众了解政府政策,也对居民的日常生活提供了实在的帮助;丹麦一位女士利用公共数据开发的网站findtoilet.dk,可以显示全丹麦公共厕所的信息,来帮助她所认识的有膀胱问题而不敢出门的人士放心外出;Zillow可以帮助用户在大量数据分析的基础上,做出购房以及居住地域的选择,类似的公司还有Ttulia、Estately、Redfin等。

大数据从最初的概念和理念演变到今天成为各国政府的行动纲领和资本市场的投资方向,单纯对大数据本身的智能分析已不足以满足政府和企业应对技术模式创新、产业垂直整合和业务深度融合的需求。面对低延迟、细粒度、多样化的数据源,如何实现复杂数据的信息协同和科学决策的智慧支撑成为新的时代命题。

1.3.2 大数据分析的关键技术

随着智慧城市发展的需求变革,大数据将最终服务于政府、企业和科研机构的科学决策,这就从信息对称和快速反应的角度对大数据本身的技术体系提出了挑战。在海量数据的技术基础上,大数据由于其自身的固有特征,在非结构化数据的海量存储与实时处理、多数据源的整合与集成、多维尺度分析、可视化分析、数据质量、安全与隐私保护等五个方面面临更大的挑战。除此之外,大数据还带来了一些其他的技术挑战,如大数据的新型表示方法、大数据的去冗降噪技术、高效率低成本的大数据存储、适合不同行业的大数据挖掘分析工具和开发环境、大幅度降低数据处理、存储和通信能耗的新技术等。

信息技术的发展经历了从硬件到软件再到服务的变迁。大数据的本质实际上是通过新一代信息化技术从各种各样的终端理解数据,快速整合,挖掘价值,并最终做出决策。

大数据的4V特征对涉及产生、管理、整合、分析、价值提取生命周期各

个环节的传统技术都带来了巨大的挑战。当前关注的大数据关键技术主要涉及海量分布式文件系统、并行计算框架、非关系型数据库(NoSQL)、实时流数据处理、内存计算及智能分析技术,如模式识别、自然语言理解、应用知识库等。大数据分析的基础技术如图 1.7 所示。

1. 开源软件受到广泛欢迎

开源项目和产品正在主导新兴的大数据市场。分布式处理的软件框架 Hadoop、用来进行数据挖掘和可视化的软件环境 R、非关系型数据库 HBase、MongoDB 和 CouchDB 等开源软件都在大数据技术领域占据重要地位。2012 年排名前 5 位的数据挖掘工具中,有 4 个是开源软件。

2. 人工智能技术不断融入

大数据可以看作是对大规模数据集合的智能分析处理。大数据之所以受到重视,是因为它能够帮助人们从似乎无穷多的数据中发现信息、发现规则、发现知识、发掘智慧,进而对未来的发展态势做出预测。要想对大数据做出这样的智能处理,就必须要用到人工智能技术,大数据的管理、分析和可视化等技术无不与人工智能相关联,目前机器学习、数据挖掘、自然语言理解、模式识别等人工智能技术已经深深融入到大数据各流程的处理技术之中。



3. 非结构化数据处理技术受到重视

云计算时代的到来使得数据创造的主体由企业逐渐转向个体,而个体所产生的绝大部分数据为图片、文档、视频等非结构化数据。信息化技术的普及使得企业更多的办公流程通过网络得以实现,由此产生的数据也以非结构化数据为主。因此,对非结构化数据的处理需求越来越强烈,非结构化处理技术越来越受到重视,非结构化数据采集技术、NoSQL 数据库、流处理技术正在快速发展。

4. 分布式处理架构成为大数据处理的普遍模式

由于大数据要处理大规模、海量、异构的数据,传统的处理方法在存储

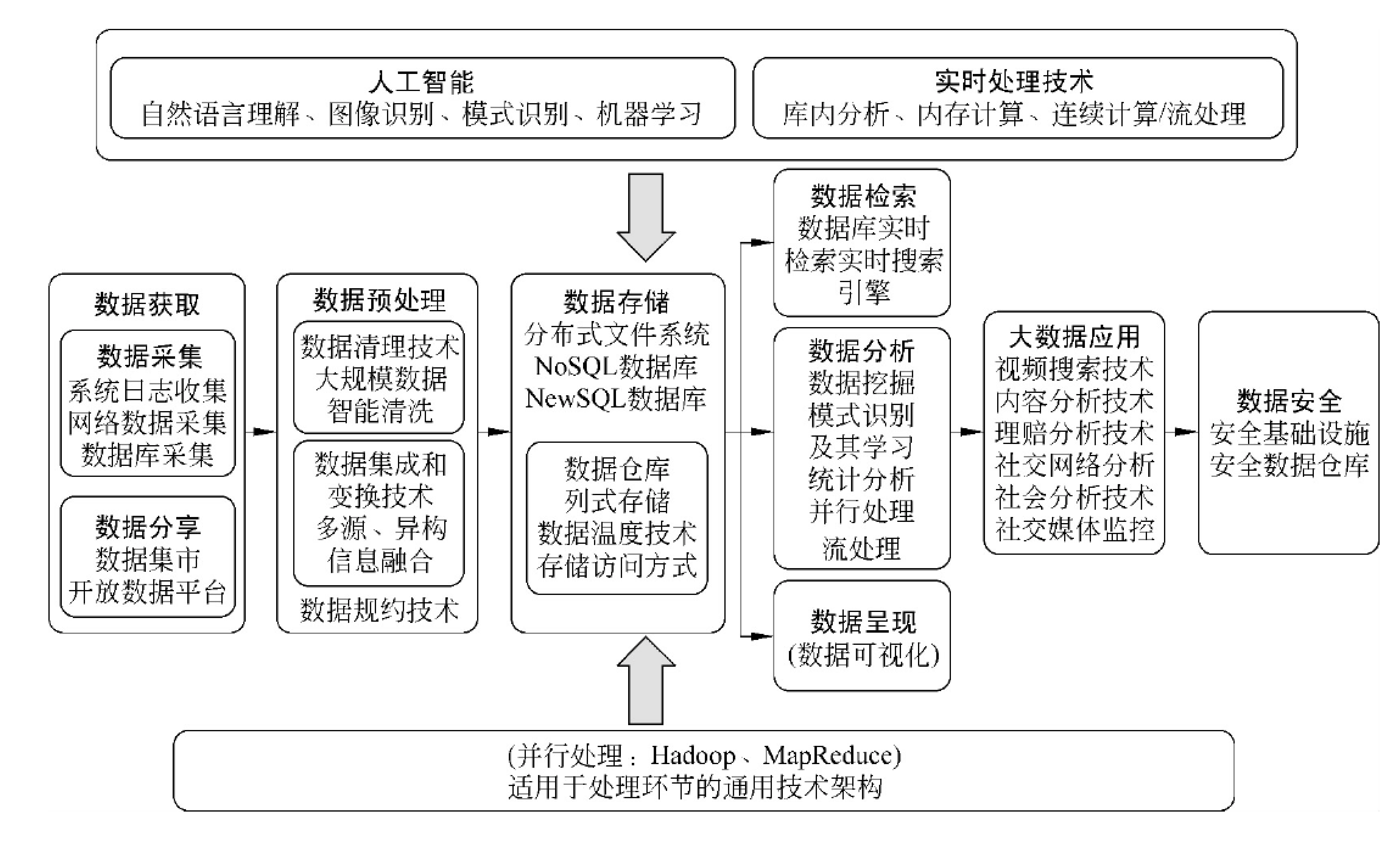


图 1.7 大数据分析的基础技术

空间、处理时间和效率上都难以满足人们对大数据处理的要求，所以在各个处理环节中都普遍采用分布式方法进行并行处理。此外，由于云计算技术的发展，利用云计算技术处理大数据问题成为人们广泛采用的方法，而云计算技术也是以分布式处理为核心的。目前，MapReduce 等分布式处理方式已经成为大数据处理各环节的通用处理方法，分布式文件系统、大规模并行处理数据库、分布式编程环境等技术也普遍被使用。

数据分析是大数据决策的核心。大数据的价值体现在对大规模数据集合的智能处理，从而可在无穷多的数据中发现信息、知识和智慧。要想实现这样的价值，最关键的步骤就是对数据的分析和挖掘。数据的采集、存储和管理都是数据分析步骤的基础，数据分析得到的智能结果可以应用到大数据相关的各个领域。大数据将充分利用机器学习、数据挖掘、模式识别、自然语言理解等人工智能基础技术，进一步实现数据分析的智能化。



2013 年 11 月，初创企业 Vicarious 开发出一种算法，可击败文字型的 CAPTCHA①。其中，被认为难度最高的 Google 的 reCAPTCHA 测试其识别率达 90%，而在 Yahoo、Paypal 及 CAPTCHA.com 的表现甚至更好，说明数据分析技术又迈出了重要一步。

能够对自然语言处理和图像识别等前沿领域提供支持的深度学习技术是大数据下最热门的趋势之一。Gigaom.com 网站整理了一个指南：深度学习领域的快速发展，鼓舞着越来越多的在自然语言处理和图像识别等领域的初创企业去研究它。同时，包括 Google、微软、Facebook 和雅虎在内的科技巨头，在深度学习方面的投入也在提高。有不少高校的研究机构也在该领域进行前沿技术的研究。深入学习技术的主要参与者如表 1.1 所示。

---

① Completely Automated Public Turing Test to Tell Computers and Humans Apart（全自动区分计算机和人类的图灵测试，CAPTCHA）：很多网站和应用都使用 CAPTCHA 来验证另一端的登录用户是否是人类。从理论上来说，CAPTCHA 可拥有多种形式，但是扭曲文字图片形式被证明是阻止恶意软件机器人程序及 SPAM 的有效方式。这是因为当文字以扭曲、重叠或被随机线、点及颜色遮盖的图片形式出现时软件很难破译；而人类这方面却能力超强，往往只需要看几眼就能识别出各种变化形式的文字。

表 1.1　深度学习技术的主要参与者

类型	公　司	关注领域	主 要 进 展
初创公司	AlchemyAPI	文本分析及图像识别	通过 API 提供服务。许多公司通过它提供的服务，提取关键词进行情感分析、内容分类和标记
	Cortica	图像识别	仿照人类在识别图像时大脑皮质中神经网络的图像处理的方式，产品出售给出版商和广告商，用以展示与页面图片内容相关的广告内容
	Ersatz	图像识别和情感分析	为深度学习设计了类似亚马逊云计算的平台产品，提供给用户网页交互界面、API、基于云端资源的 GPU 和神经网络的实现，能够让用户以需要的方式，组建和运行模型
	Semantria		通过 API 和 Excel 插件实现功能，通过整合更多深度学习的方法和扩展除维基百科（Lexalytics 引擎获取语义知识的地方）之外的数据源，来提高服务的精度
科技巨头	Facebook		希望深入学习可更好地优化 News Feed 的功能，并实现更畅快的照片共享体验。Facebook 对于深入学习领域最大的贡献可能是它数量众多的基础设施和开源的专业硬件
	Google	文本、图片、语音	Google 是深入学习领域中最著名的公司，这要归功于它高度公开的图像识别领域的研究（这个模型不需要训练，就能够识别猫脸和人脸）。最近它又决定开源一些文本分析的新工具。深度学习已经支持 Andriod 手机上的语音识别，还能直接在 Google＋上搜索图片，即便这些图片没有任何标签
科技巨头	IBM		Watson 系统在智力竞赛节目《危险边缘》中，击败了所有对手获得冠军。现在，这个技术被应用到医疗保健等领域，它整合了大量的数据分析技术，其中就包含深度学习。之外，IBM 还围绕“认知计算”做了大量的工作。IBM 联合四所高校进行科研合作，并把深度学习作为其中的一个核心
	微软	文本、图片、语音	微软研究院对大数据进行了多年的研究。其研究人员想方设法从网络和移动应用收集各种数据，并希望深入学习能在网页、移动、游戏，甚至商业软件平台等方面提供更加富有魅力的体验，而这背后最大的王牌可能是 Kinect 技术



续表

类型	公　司	关注领域	主 要 进 展
科技巨头	雅虎	图像识别	雅虎没有像 Google 和微软那样引人注目，但它分别收购了两家基于深度学习的图像识别公司 IQ Engines 和 LookFlow。雅虎目前的重心是让 Flickr 变得更具吸引力，使其在移动端的设置更加简单、智能
研究机构	多伦多大学研究小组	语义分析	在深度学习领域有许多重大的突破，2013 年创建的 DNNresearch 已被 Google 收购
	斯坦福大学研究小组		斯坦福大学的研究侧重于对整个句子的理解，而不是单个单词。目前，对于分析电影评论情感分析的准确率已经达到 85%
	纽约大学研究小组	图像识别	研究让计算神经网络更加简单
	蒙特利尔研究小组		蒙特利尔大学的 LIST 实验室创造了开源数据库 Theano。它能使得复杂的程序设计语言 Python 变得更加简单，同时可让这种语言在 GPU 上运行
	瑞士研究小组		1991 年就开始研究深度学习。该团队已经赢得了无数深度学习领域的竞赛



除此之外，大数据还带来了一些其他的技术挑战，如大数据的新型表示方法、大数据的可视化分析、大数据的去冗降噪技术、高效率低成本的大数据存储、适合不同行业的大数据挖掘分析工具和开发环境以及大幅度降低数据处理、存储和通信能耗的新技术等。

1.3.3　大数据的安全和隐私保护

大数据对数据的完整性和可用性带来了挑战，但在防止数据丢失、被窃取和被破坏上存在一定的技术难度，传统的安全工具不再像以前那么有效，而且大数据技术也可能成为黑客的攻击手段和载体。

1. 大数据使个人隐私保护更为困难

20 世纪 90 年代，可以通过“性别＋邮编＋出生年月日”识别出 87%的美国人，只要保护好这些个人信息就能很好地保护隐私。但在今天，通过分析

用户 4 个曾经到过的位置点就可以识别 95% 的用户，分析 Facebook 的 like 按钮就能获得大量用户个人信息，如种族（准确率 95%）、性别（93%）、性取向（88%）、单身状况（67%）等，这使得保护个人隐私更为困难。

2. 国家安全面临更大威胁

2013 年 5 月底，随着“斯诺登事件”爆料棱镜（Prism）计划，美国国家安全局（NSA）秘密对其国内公民和其他国家的大规模数据收集和监控行为浮出水面，谷歌、微软等 6 家互联网企业和运营商为其提供了数据支持。其中，无界告密者（Boundless Informant）项目在 2013 年 2 月至 3 月的短短 30 天内，就从全世界互联网上收集到 970 亿条数据。据美国媒体披露，NSA 还有一个名为定制入口组织（Tailored Access Operations）的秘密机构，有超过 1000 名军队及民间专家，该组织每小时可获取 2PB 数据并可自动处理。英国也有类似的大数据监控行动，并与美国共享情报。中国大陆是美国监控的重点对象。据斯诺登称，在过去 15 年时间里，定制入口组织已渗透到中国计算机及电信系统，获得了有关中国“最好的、最可靠的情报”。

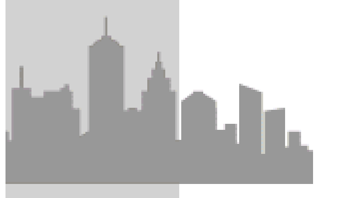
3. 数据安全的风险更加凸显

随着大数据海量数据存储和信息分析萃取手段的进步，必将加大信息的开放度，带来的副作用是 IT 基础架构将变得越来越一体化和外向型，这就对数据安全和知识产权构成了更大的风险。同时，由于大数据拓宽了对个人信息获取的渠道，引发了用户隐私性与信息利用便利性之间的冲突。在大数据时代，不论告知与许可，还是匿名（隐身）、模糊化，这三大传统的用户隐私保护策略都将失效。消费者虽然可以受惠于海量数据挖掘带来的更符合消费需要和更低价格的商品，但也随着个人购买偏好、健康、财务情况等数据被收集，增加了对隐私安全的担忧。因此，必须采取应用和管理同步、开放和管控并重的方法，在促进大数据时代市场良性竞争的同时，切实保护数据安全。

随着数量越来越多的数据被数字化，跨越组织边界而流动着，一系列政

策问题将会变得越来越重要，这包括但不限于隐私、安全、知识产权和责任。显然，随着海量数据的价值愈加明显，隐私是个重要等级（尤其是对消费者来说）不断提高的问题。个人数据（例如健康和财务记录）经常能够提供最重要的人类福利，例如，帮助精准确定适当的医疗或者最恰当的金融产品。然而，消费者也将这些类别的数据视为最敏感的个人隐私。显然，个人和其生活所在的社会将不得不努力在数据隐私和数据的功用之间权衡取舍。

海量数据日益提升的经济意义也昭示了一系列法律问题，尤其是当其与如下事实联系起来时：即数据与许多其他资产具有根本性的差异。数据可以与其他数据结合起来完美而轻松地复制，同样一份数据可以由多个人同时使用。这些是数据与实体资产相比的独有特征。有关数据所附带的知识产权的问题不容回避：何人“拥有”某份数据？某一数据集附带着何种权利？数据的“公平使用”的定义是什么？此外，还有与责任相关的问题：当一份不准确的数据导致负面结果时谁应负责？要充分发挥海量数据的潜力，此类法律问题需要澄清，也许会随着时间的推移逐步澄清。



4. 大数据跨境流动问题更加频繁

跨境数据服务折射出隐私安全。2011 年，两位美国技术人员发现 iPhone 会在隐蔽的情况下持续收集用户位置信息并且保存。《华尔街日报》调查后发现，即便用户关闭手机的定位功能，也不能阻止这种情况发生。多数用户在使用手机应用商店服务时，都会“同意”所谓的隐私协定，但他们并不清楚这个简单的行为意味着什么。这些应用商店和软件开发者会利用这些资料去做用户完全不知道的事情。无论用何种方式登录互联网，许多隐藏在背后的机构和个人可以瞬间知道你是谁、来自哪里、收入高低、品牌喜好，甚至一周内的消费计划。而孤立的用户永远不会想到，这些信息就是通过 iPhone 的一次不经意下载、搜索、导航、点评被掌握的。

目前，我国对大数据的安全保护能力还十分有限。当前，发达国家已经借助大数据发展逐步掌握窃取、挖掘别国信息的能力，“棱镜门”事件就是这

一现象的集中反映。我国对大数据的保护能力还十分有限，数据被恶意使用的现象难以掌控。2012 年中国的数据存储量是 364EB，其中 55%（200EB）的数据需要一定程度的保护，然而只有 96EB 的数据得到保护；个人和企业的数据资源保护意识比较单薄，如 CSDN 600 万个人注册信息泄露，中国人寿 80 万保单个人信息泄露。

1.3.4 大数据决策的应用案例

目前，大数据决策正在向更多的行业和应用场景拓展。在行业方面，大数据决策已经从电子商务、互联网、快速消费品等行业向金融、政府、公共事业、能源、交通等行业扩展；从应用场景来看，也从结构化数据的分析发展到半结构化、非结构化数据的分析，尤其是社交媒体信息分析受到用户的更多关注。

1. 谷歌

大数据决策最著名的案例当属 2009 年甲型 H1N1 流感爆发几周前，互联网巨头谷歌公司的工程师们在《自然》杂志上发表了一篇引人注目的论文。它令公共卫生官员们和计算机科学家们感到震惊。文中表示，和疾控中心一样，谷歌也能判断出流感是从哪里传播出来的，而且其判断非常及时，不会像疾控中心一样要在流感爆发一两周之后才可以做到。谷歌公司发现能够通过人们在网上检索的词条辨别出其是否感染了流感后，把 5000 万条美国人最频繁检索的词条和美国疾控中心在 2003—2008 年间季节性流感传播时期的数据进行了比较。通过一个数学模型处理后，其预测与官方数据的相关性高达 97%。所以，2009 年甲型 H1N1 流感爆发的时候，与习惯性滞后的官方数据相比，谷歌成了更有效、更及时的指示标，公共卫生机构的官员获得了非常有价值的数据信息。

2. 零售行业

诸如沃尔玛、Tesco（英国零售巨头）等巨头已从数据中获得了巨大的利

益，也因此巩固了自己在业界的长盛不衰。以曾经因“少女怀孕事件”而成为大数据典型案例的 Tesco 公司为例，这家全球利润第二大的零售商从其会员卡的用户购买记录中，可充分了解一个用户是什么“类别”的客人，如速食者、单身、有上学孩子的家庭等，并可基于这些分类进行一系列业务活动。比如，通过邮件或信件寄给用户的促销可以变得十分个性化，店内的上架商品及促销也可以根据周围人群的喜好、消费的时段使其更加有针对性，从而提高货品的流通。这样的做法为 Tesco 获得了丰厚的回报，仅在市场宣传一项，就能帮助 Tesco 每年节省 3.5 亿英镑的费用。



3. 能源行业

SaaS 型软件公司 Opower 使用数据来提高消费用电的能效，并取得了显著的成功。Opower 与多家电力公司合作，分析美国家庭的用电费用并将其与周围邻居的用电情况进行对比，被服务的家庭每个月都会收到一份对比的报告，显示自家用电在整个区域或全美类似家庭中所处的水平，以鼓励节约用电。据报道，Opower 的服务已覆盖了美国几百万户居民家庭，预计可为美国消费用电每年节省 5 亿美元。

4. 生物信息行业

生物信息是继互联网行业之后数据迸发最迅速的行业，并将远远超过互联网产生的数据：人类用 0 和 1 创造了虚拟世界，而大自然用 A、C、T、G（腺嘌呤 Adenine、胞嘧啶 Cytosine、胸腺嘧啶 Thymine、鸟嘌呤 Guanine）四种元素创造了万物，生命的产生、发展、消亡的奥秘尽在其中。随着测序技术的发展，全基因组的测序价格由十年前的上亿美元降至今天的数千美元，这使得更多人、物种的 DNA 信息的获取成为可能。个体全基因组信息的获取，使得个性化诊疗服务成为可能。大数据时代，一切的一切都存在着可能，而这一切的改变我们也正在体验之中。

5. 阿里巴巴

阿里巴巴旗下的淘宝网是全球访问量最大的电子商务网站。以前，淘

宝网的事务处理依托国际数据库巨头甲骨文的商业数据库软件,成本很高,但性能和可扩展性较差,制约了企业发展。几年前,淘宝网下决心使用开源软件 MySQL Cluster 替代,其事务处理的性能和可扩展性都有了数量级的提升。现在淘宝网的全部应用软件(包括数据魔方等数十种大数据计算应用)和基础软件都是自行开发或采用开源软件,摆脱了商业软件制约。阿里公司根据在淘宝网上中小企业的交易状况筛选出财务健康和讲究诚信的企业,对其发放无须担保的贷款。目前已放贷 300 多亿元,坏账率仅 0.3%,大大低于商业银行。

淘宝网还建立了“淘宝 CPI”,通过采集、编制淘宝网上 390 个类目的热门商品价格来统计 CPI,比国家统计局公布的 CPI 还提前半个月预测经济的走势。

6. 华大基因公司

华大基因是目前世界上最大的基因组学研究中心,也是我国典型的大数据处理和应用公司。该公司建立了大规模基因测序、克隆、农作物基因组等技术平台,测序和基因组分析能力世界领先。目前,华大基因已经完成了水稻、谷子、玉米、大豆、番茄等重要农作物的全基因组测序,对 25 种栽培稻和 24 种野生稻进行了基因组扫描和分析,找到了 162 个基因,这些基因与水稻高产性状紧密相关。该公司还启动了百万人基因图谱计划,预计 3~5 年内测定 100 万人以上的全基因组图谱,目前已针对染色体疾病等多种疾病开发了先进的基因检测技术,形成了贯穿整个生命周期的基因检测与诊断技术体系。

7. 农夫山泉

农夫山泉通过大数据分析技术使销售额提升了大约 30%,并使库存周转从 5 天缩短到 3 天,同时其数据中心的能耗降低了约 80%。

8. 百分点公司

百分点公司拥有海量网购消费者偏好数据,积累了超过 1.4 亿名网购消

费者的消费偏好和200多亿个消费偏好标签，已成为国内最大的跨网站消费偏好平台。

9. 中信银行

中信银行信用卡中心通过部署大数据分析系统，实现了近似实时的商业智能和秒级营销，运营效率得到全面提升，每次营销活动配置平均时间从2周缩短到2～3天，交易量增加65％，不良贷款比率同比减少了0.76％。

第 2 章　开放数据与政府信息资源利用

2.1　信息资源管理的基本框架和关键技术

2.1.1　传统信息资源管理的基本框架

信息资源管理主要包括信息产生、信息采集、信息处理、信息开发利用和信息服务五个阶段（见图 2.1）。

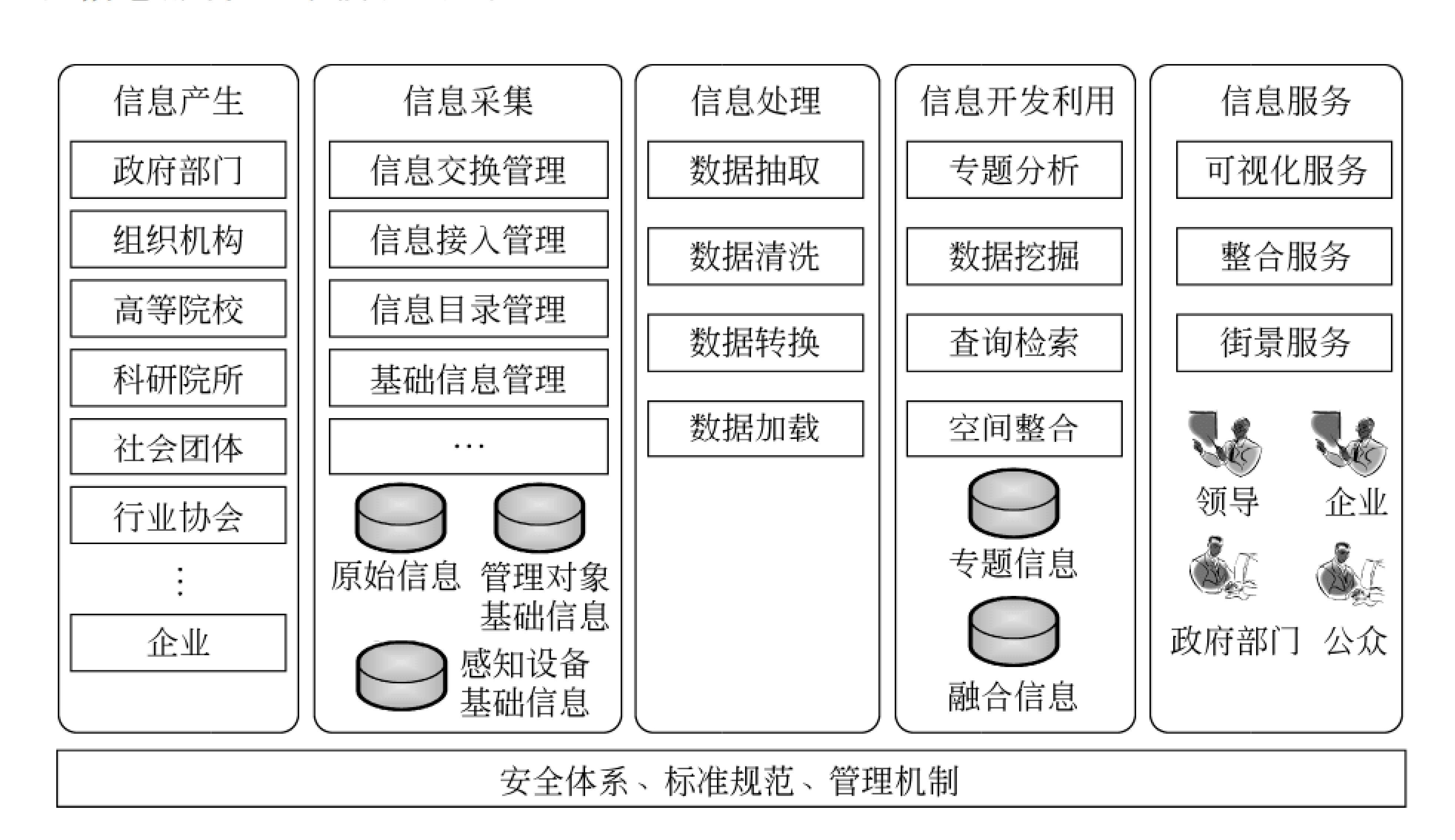


图 2.1　信息资源管理的基本框架

1. 信息采集

信息的采集方式主要有四种。

（1）手工填报：主要用于层级管理中，由下层部门通过信息系统中的相关流程进行逐项填写后上报，或采用导入的方式批量填报。

（2）数据获取：主要通过部署在物理世界中的感知设备进行实时感知信息采集，感知信息通过集中器、终端、传输网络通道等进行传输汇聚。



（3）数据交换：主要通过在各信息源部署前置机的方式，实现多信息源间的数据互联互通。

（4）数据接口：主要通过数据接口调用的方式实现数据共享和服务。

采集内容主要分为五类。

（1）基础信息：主要指管理对象、感知设备等实体的基础属性信息。

（2）实时信息：主要指来源于感知设备的实时感知信息。

（3）交换信息：主要指由相关的信息所有部门对采集的数据进行解析处理后，按照一定的交换规则共享的信息。

（4）综合信息：主要指由相关的信息所有部门将实时感知信息与基础信息和业务领域的主题信息进行整合融合后的信息。

（5）资源描述信息：主要指相关的信息所有部门可提供共享的各类共享服务资源的元数据，如应用系统元数据、Portlet 元数据、页面元数据、数据库元数据、Web 服务元数据等。



2. 信息处理

信息处理主要完成对采集信息的加工整理，对采集过程中可能出现的噪声数据进行清洗，转换成有效数据。

（1）数据抽取：从数据采集过程形成的临时库、基础库中抽取相关的数据。

（2）数据清洗：清除脏数据（dirty data）或噪声，以保证数据质量。

（3）数据转换：通常不仅仅是数据格式的转换，外部系统中的数据可能包含不一致或者不正确的信息，这取决于外部系统中的数据情况。

转换步骤的一部分是“净化”或“拒绝”不符合条件的数据，这个阶段常用的技术包括字符检查（拒绝包含字符的数值性数据）和范围检查（拒绝超出可接受范围的数据）。被拒绝的记录通常存放在单独的文件中，使用更复杂的工具处理（或者手工改正问题），然后将这些数据合并到已转换的集合中。主要的转换方式包括以下五种：一是数据类型转换，将数据源中的不同

数据类型转换成需要的数据类型；二是数据表示方法转换，命名代码统一，汉字编码同义，度量衡表示统一以及其他数据表示方法的统一等；三是命名转换，将数据模式、表名、属性名转换成统一命名方式；四是数据综合，按粒度要求对动态属性数据进行统计汇总形成综合性数据；五是数据筛选，按照分析及决策的需要从数据源中作纵向的属性选择及横向的实例选择。

（4）数据加载与刷新：将获取并转换的数据存放到新的数据库中。

3. 信息资源开发利用

信息资源的开发利用主要是按照需求将采集到的信息进行整合，为上层应用等提供支撑，主要包括以下三个方面。

（1）数据关系映射：将管理对象、感知设备、实时信息、信息主体（部门）等之间建立关联关系，形成支持应用的主题信息库，在主题信息库的基础上形成信息融合库。

（2）分析建模：按照应用需求，建立分析模型，为领域应用提供调用服务。

（3）空间信息整合：通过物联信息、决策模型与空间图层的整合，实现信息的可视化、全景化、空间化呈现。

4. 标准规范

标准规范的重点主要包括多源信息的统一编码、基础信息的属性规范、多源信息的共享流程、多源信息的交换与传输、信息目录、信息接入方式等。

5. 信息管理

信息管理主要包括数据更新管理、数据使用管理以及数据的存储和备份。在数据更新时，必须进行数据质量控制。对入库数据质量进行严格把关。在数据入库时，对数据的合法性进行检查，并对数据之间的关系建立关联，检查源及目的数据结构的逻辑对应关系是否正确；审核数据更新标志；然后在数据提交过程中检查数据及代码的完整性、合法性，保证数据一致性。



（1）更新管理：系统数据要具有现势性，即数据要反应最新的现实情况。因此要建立和形成一种有效的、实时的数据更新机制，同时发展实用的、经济快捷的更新手段和技术方法，以保持数据的现势性，这样才能保证系统进行的查询、分析、咨询决策等结果的正确性。

（2）日志记录：数据处理日志分为三类。一是数据处理执行过程日志，是在数据处理执行过程中每执行一步的记录，用流水账形式记录每次运行每一步骤的起始时间及影响了多少行数据；二是异常日志，当某个执行过程出错的时候写错误日志，记录每次出错的时间、出错的位置以及出错的信息等；三是任务日志，只记录任务开始时间、结束时间及是否成功等信息。

（3）数据库系统安全访问控制：严格控制数据库系统的访问控制权限，对不同的用户进行不同数据库的访问控制，同时删除不用的数据库用户，确保非法用户对数据库系统的访问；对数据库用户的密码严格保密，使其不被不相关的人员非法获取；为数据库系统设置防火墙，将数据库系统设置到防火墙内，利用防火墙的安全访问控制策略，分别控制不同的用户、IP 对数据库系统的访问级别；限制数据库系统的客户端对数据库系统的非法访问。



（4）备份与恢复：为了确保数据安全可靠，使信息系统正常运行，需要对信息系统的数据进行定期备份，以便在发生数据库严重故障时利用备份的数据进行恢复。数据的备份策略根据不同的数据进行不同的定义。初始化数据在加载完成后进行一次性的备份；配置数据、控制数据在信息系统每次配置变动后立即进行自动备份；信息系统的业务数据需要建立详细的备份策略实现联机和脱机两种备份。

2.1.2 物联信息管理的关键技术

随着摄像头、传感器等各种感知终端的普及应用，来源于物理世界的物联信息日益庞大，在信息源、信息载体、传输频率等方面具有区别于传统信息资源的鲜明特征，正逐渐成为城市大数据的主体。从物联网的技术体系上看，物联信息的管理涉及到“感、传、知、用”四个层面的关键技术。

1. 感知层关键技术

感知层技术是物联信息采集的核心技术，按照国际电信联盟（ITU）的划分，感知层的关键技术主要包括射频识别技术（RFID）、传感器技术、智能嵌入技术等。

1）射频识别技术

射频识别技术是20世纪90年代兴起的一种非接触式自动识别技术，该技术的商用促进了物联网的发展。它通过射频信号等一些先进手段自动识别目标对象并获取相关数据，有利于人们在不同状态下对各类物体进行识别与管理。

射频识别系统通常由电子标签和阅读器组成。电子标签内存有一定格式的标识物体信息的电子数据，是代替条形码走进物联网时代的关键技术之一。该技术具有一定的优势：能够轻易嵌入或附着，并对所附着的物体进行追踪定位，读取距离更远，存取数据时间更短；标签的数据存取有密码保护，安全性更高。RFID目前有很多频段，其中集中在13.56MHz频段和900MHz频段的无源射频识别标签应用最为常见。短距离应用方面通常采用13.56MHz HF频段；而900MHz频段多用于远距离识别，如车辆管理、产品防伪等领域。阅读器与电子标签可按通信协议互传信息，即阅读器向电子标签发送命令，电子标签根据命令将内存的标识性数据回传给阅读器。

RFID技术与互联网、通信等技术相结合，可实现全球范围内物品跟踪与信息共享。但其技术发展过程中也遇到了一些问题，主要在于芯片成本；还有RFID反碰撞防冲突、RFID天线研究、工作频率的选择及安全隐私等问题，也在一定程度上制约了该技术的发展。

2）传感器技术

国家标准（GB/T 7665—2005）中对传感器的定义是：能感受被测量并按照一定的规律转换成可用输出信号的器件或装置，通常由敏感元件和转换元件组成。传感器技术是涉及物理学、化学、生物学、材料科学、电子学以

及通信与网络技术等多学科交叉的高新技术，作为当代信息技术中信息获取的重要手段，已广泛应用于工业、农业、国防、医疗卫生等人民生活和国民经济建设的各个领域。

传感器技术与通信技术、计算机技术一起被称为信息技术的三大支柱，分别对应了“感”、“传”、“知”三个层面。传感器技术是从自然信源获取信息，并对之进行处理(变换)和识别的多学科交叉的现代科学与工程技术。传感器技术的核心即传感器，它是负责实现物联网中物与物、物与人信息交互的重要组成部分。

目前，传感器技术已由传统的机械结构型，经过机电型、固体传感器、集成传感器进入到微传感器和微系统的阶段，并朝着微型化、集成化、智能化、网络化的方向发展。

按照传感器的监测数据的不同，可以将传感器划分为三种：一是应用于工业领域的传感器，如温度传感器、压力传感器、物理量传感器、流量传感器等；二是民用领域传感器，如 RFID 传感器、二维码光学传感器等；三是多媒体类型的传感器，如音频传感器、视频传感器、无线音频视频传感器、可自由编程万能音频传感器等。



目前传感器技术越来越受到普遍的重视，它的应用已渗透到国民经济的各个领域，涵盖工业生产过程的测量与控制、汽车电控系统、现代医学、环境及军事等多个方面。大规模集成电路、微纳加工、网络等技术的发展，为传感器技术的发展奠定了基础。随着生产自动化程度及人们生活水平的日益提高，对传感器的要求也日益提高。技术推动和需求牵引共同决定了未来传感技术的发展趋势，主要包括四个方面。

(1) 微型化：采用微加工手段和纳米制备技术，可加工出特征尺寸达到微米甚至纳米的微型器件，同时带来功耗的降低。

(2) 集成化：包括传感器与 IC 的集成制造技术以及多参量传感器的集成制造技术，缩小了传感器的体积，提高了检测精度。

(3) 智能化：智能化是在集成化基础上的更进一步发展，使得信号检测

具有一定的智能，包括传感器的自校准，可根据被测量的变化实时调节量程和精度、模式识别等。

(4) 网络化：传感器网络化极大地增强了传感器的探测能力，是近几年来的新的发展方向。

3）嵌入式系统技术

嵌入式系统是以应用为中心，以计算机技术为基础，其软硬件可裁剪，适用于对功能、可靠性、成本、体积、功耗有严格要求的专用计算机系统。它一般由嵌入式微处理器、外围硬件设备、嵌入式操作系统以及用户的应用程序四个部分组成，具有对其他设备的控制、监视或管理等功能。

目前，大多数嵌入式系统还处于单独应用的阶段，以微控制器(Microcontroller Unit，MCU)为核心，与一些监测、伺服、指示设备配合实现一定的功能。互联网现已成为社会重要的基础信息设施之一，是信息流通的重要渠道。如果嵌入式系统能够连接到互联网上，则几乎可以方便、低廉地将信息传送到世界上的任何一个地方。



2. 传输层关键技术

传输层主要负责信息传递和信息控制，提供端到端的交换数据的机制，实现物与物之间的“互联”。目前物联信息传输的关键技术主要包括 3G 技术、4G 技术、蓝牙技术、智能网关技术(NFC)等。

1）3G 技术

第三代移动通信技术(3rd-generation，3G)是指支持高速数据传输的蜂窝移动通信技术。3G 服务能够同时传送声音及数据信息，速率一般在几百 kbps 以上。3G 主要存在四种标准：CDMA2000、WCDMA、TD-SCDMA、WiMax。第三代移动通信系统的一个突出特色就是：在未来移动通信系统中实现个人终端用户能够在全球范围内的任何时间、任何地点，与任何人，用任意方式，高质量地完成任何信息之间的移动通信与传输。

2）4G 技术

4G 技术又称 IMT-Advanced 技术。准 4G 标准是业内对 TD 技术向 4G 发展的 TD-LTE-Advanced 的称谓。

4G 通信系统采用了一些不同于 3G 的技术。4G 中将使用的核心技术，总结起来，有下列几种：正交频分复用（Orthogonal Frequency Division Multiplexing，OFDM）、软件无线电、智能天线技术、多输入多输出（Multiple-Input Multiple-Output，MIMO）、基于 IP 的核心网。国际电信联盟（ITU）划定的 4G 标准主要有五种：LTE-Advanced、WirelessMAN-Advaneed、WiMax、HSPA+和 LTE。

3）蓝牙技术



蓝牙技术是一种支持设备短距离通信（一般 10m 内）的无线电技术。它能在包括移动电话、PDA、无线耳机、笔记本电脑、相关外设等众多设备之间进行无线信息交换。利用蓝牙技术，能够有效地简化移动通信终端设备之间的通信，也能够成功地简化设备与 Internet 之间的通信，从而使数据传输更加迅速、高效，为无线通信拓宽道路。蓝牙技术采用分散式网络结构以及快跳频和短包技术，支持点对点及点对多点通信，工作在全球通用的 2.4GHz ISM（即工业、科学、医学）频段，其数据速率为 1Mbps，采用时分双工传输方案实现全双工传输。

4）智能网关技术

智能网关技术是应用网关技术的一种，其主要功能是自动完成对系统中大量基站监控数据的提取、处理和转发，实现系统之间的交互连接与对话。目前，智能网关技术广泛应用于通信、移动、家庭等各个方面。

物联网网关可以实现感知网络与通信网络，以及不同类型感知网络之间的协议转换，既可以实现广域互联，也可以实现局域互联。此外物联网网关还需要具备设备管理功能。运营商通过物联网网关设备可以管理底层的各感知节点，了解各节点的相关信息，并实现远程控制。

3. 智能处理层技术

智能处理层综合运用高性能计算、人工智能、数据库和模糊计算等技术,对收集的感知数据进行通用处理,主要涉及海量数据存储技术、云计算技术、数据挖掘技术、SOA 技术、中间件技术等关键技术。

1）海量数据存储技术

海量信息存储是一个包括网络设备、存储设备、服务器、应用软件、公共访问接口、接入网络和客户端程序等多个组成部分的系统。基本是以服务器为中心的处理模式,使用直连存储(Direct Attached Storage,DAS),存储设备(包括磁盘阵列、磁带库、光盘库、硬盘等)作为服务器的外设使用。

随着网络技术的发展,服务器之间交换数据或向磁盘库等存储设备备份时,都是通过局域网进行,主要应用网络附加存储(Network Attached Storage,NAS)技术来实现网络存储,将占用大量的网络开销,严重影响网络的整体性能。为了能够共享大容量的高速度存储设备,不占用局域网资源进行海量信息传输和备份,通常需要专用存储网络来实现。

2）云计算技术

云计算(Cloud Computing)是分布式计算技术的一种,通过网络将庞大的计算处理程序自动分拆成多个较小的子程序,再交由多部服务器所组成的庞大系统经搜寻、计算、分析之后将处理结果回传给用户。云计算的核心内涵是计算服务化、资源虚拟化和管理智能化。云计算的核心思想是将大量用网络连接的计算资源统一管理和调度,构成一个计算资源池向用户提供按需服务。

云计算系统的关键技术主要包括编程模型、数据管理技术、数据存储技术、虚拟化技术、云计算平台管理技术等。

3）数据挖掘技术

数据挖掘(Data Mining,DM)是从存放在数据库、数据仓库或其他信息库的大量数据中获取有效的、新颖的、潜在有用的、最终可理解的模式的非



平凡过程。在人工智能领域,数据挖掘习惯上又称为数据库中的知识发现(Knowledge Discovery in Database, KDD),数据挖掘也是数据库中知识发现过程的一个基本步骤。

现在,数据挖掘技术已经发展成熟的三种基础技术是海量数据搜集、强大的多处理器计算机和数据挖掘算法,并已经广泛应用于商业数据仓库和计算机自动收集的数据记录等超大规模数据库。

数据挖掘的工作过程主要包括数据的抽取、数据的存储和管理、数据的展现等。数据挖掘的工作过程如图 2.2 所示。

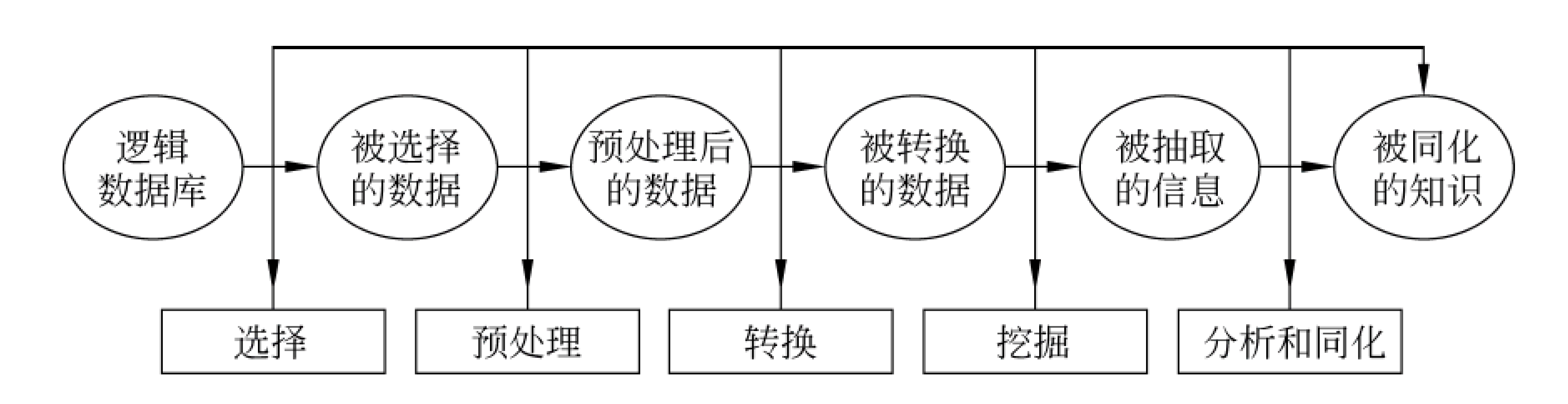


图 2.2　数据挖掘的工作过程

4）SOA 技术

SOA(Service-oriented Architecture,面向服务的体系架构)技术是一种松耦合的软件组件技术,它将应用程序的不同功能模块化,并通过标准化的接口和调用方式联系起来,实现快速可重用的系统开发和部署。SOA 可提高物联网架构的扩展性,提升应用开发效率,充分整合和复用信息资源。

5）中间件技术

中间件是一种独立的系统软件或服务程序。分布式应用软件借助中间件技术在不同的技术之间共享资源。通过中间件相连接的系统或应用程序可以工作于多平台或操作系统(Operating System,OS)环境,并实现信息的高效交换。

4. 应用层关键技术

物联网应用层的相关技术主要包括家庭物联网应用涵盖的技术和企业物联网应用的相关技术。家庭物联网应用的相关技术比如家电智能控制技

术、家庭安防监控技术。企业物联网应用的相关技术现阶段比较典型的技术主要包括智能家电控制技术、石油监控应用技术、电力抄表、车载应用技术等。此外，还有对这些应用进行支撑的网络管理技术和安全保障技术。

1）智能家电控制技术

智能家电是微处理器和计算机技术引入家电设备后形成的家电产品，是具有自动监测、自动测量、自动控制及自动调节与远方控制中心通信功能的家电设备，可通过物联网的相关通信协议和控制设备相连并进行通信。智能家用电器目前所采用的智能控制技术主要是模糊控制等技术。少数高档次的家用电器也用到神经网络技术（也叫神经网络模糊控制技术），模糊控制技术目前是智能家用电器使用最广泛的智能控制技术。原因在于这种技术和人的思维具有一致性，可以用相对简单的数理逻辑在单片机或嵌入式设备上进行构造。

2）家庭安防监控技术

基于物联网的家庭安防监控技术区别于传统家庭安防监控技术的地方主要是采用RFID、传感器以及M2M（Machine to Machine，机-机）设备来完成家庭安防的监控，基于物联网的家庭安防系统主要由主控模块、图像采集模块、无线通信模块和传感器模块组成。其中无线通信模块通过ZigBee无线网络与传感器模块相连，完成家庭安防的监控数据采集。常用监控模式通过传感器模块采集突发的环境参数时，图像采集模块开始采集家庭实时图像；主控模块将图像发送到户主手机，户主确认是否有意外发生，然后在远程做出相应处理。基于物联网的家庭安防监控技术具有安装方便、成本低、人性化、操作方便、安全性高、有效安防等特点。

3）石油监控应用技术

基于物联网的石油监控应用技术是指通过传感器等相关物联网设备完成石油存储库区的烟、火、温度、湿度等环境因素的采集，通过网络将各个传感器点的数据进行传送和远程集中，并完成远程监控。当烟、火、温度等环境因素发生异常时，能及时反映到监控人员或监控设备，并可与摄像监控设

备相结合，进行石油库区监控的防火防盗。

4）电力抄表技术

电力抄表技术通常采用在家庭用户的电表设备上安装传感器或其他通信装置，通过电力网络的通信模块与之通信，完成家庭用户用电数据的远程抄取，完成电力设备的智能化、远程化、自动化管理。

5）车载应用技术

基于物联网的车载应用技术主要通过无线设备采集车载物联网设备的信号，可以实时获得车辆的相关位置、速度、行驶方向等信息，并可通过相关语音通信协议建立与车辆的实时语音通信，基于物联网技术的车载应用技术目前有较广阔的发展前景。目前较常用的车载应用之一是通过车载RFID或传感器实时获得车辆的位置信息，并完成GPS信息的实时上传和下载。

6）网络管理技术



物联网具有“自治、开放、多样”的自然特性，这些自然特性与网络运行管理的基本需求存在着突出的矛盾，需研究新的物联网管理模型与关键技术，保证网络系统正常高效地运行。

7）安全保障技术

安全是基于网络的各种系统运行的重要基础之一，物联网的开放性、包容性和匿名性也决定了不可避免地存在信息安全隐患，需要研究物联网安全关键技术，满足机密性、真实性、完整性、抗抵赖性的四大要求，同时还需解决好物联网中的用户隐私保护与信任管理问题。

2.2 开放数据：传统信息壁垒的破局点

开放数据运动已在全球逐步兴起，各国政府、主要城市和地区都已经意识到开放的数据是一个具有巨大潜力的未开发的资源，是一个国家或地区的重要资产。从国家和区域的层面上看，进行合理的统筹规划，整合地方和

部门的不同信息源，建设统一的数据开放门户，围绕社会需求逐步向公众开放免费、机器可读的数据集，鼓励第三方的机构或个人基于数据集开发各种应用程序，服务于政府管理、企业经营和大众生活，已成为大势所趋。

221 全球开放数据运动

根据英国开放知识基金会(Open Knowledge Foundation)的定义，"开放"(openness)需具备以下三项基本元素。

(1) 非歧视性：数据若开放，则其对任何人都开放。

(2) 机器可读性：数据若开放，则应是机器可读格式。例如对于表格数据，应该采用.csv，而非.pdf。

(3) 开放授权性：数据若开放，则其对应授权条款应确保使用者自由免费访问、获取、使用、加值、演绎、复制、传播的权利。

随着互联网、移动互联网等的持续发展，政府、企业、行业组织和个人等都收集了广泛的、不同类型的数据。但是，人们对于各种经过官方处理之后的统计数据普遍持有一定的怀疑态度；而未处理的原始数据，大部分人又难以理解。人们逐渐意识到，这些庞大数据资源的妥善开放利用，将产生巨大的社会价值和经济价值。2010 年 4 月，互联网之父蒂姆·伯纳斯·李在 TED 大会[①]上号召大家把公共数据或私人数据共享，使更多的人能够加以利用，创造出有用的甚至是令人意想不到的结果，由此开启了"开放数据运动"。

开放数据是一种新的哲学理念与实践，它按照用户特定的需求和相应的协议、规则、框架，对数据进行采集、存储、发布、加工、链接等，以实现局域或全部数据可以被任何人自由获取、互通、共享、重用，没有来自版权、专利或其他机制的限制。

---

① TED 是 Technology(技术)、Entertainment(娱乐)、Design(设计)的缩写，是美国的一家私有非营利机构的名称。该机构以它组织的 TED 大会著称，这个会议的宗旨是"用思想的力量来改变世界"。



一般而言，开放数据具有三个典型特征：一是可获取性和可访问性；二是数据应当使用允许再利用和再分发的许可协议；三是普遍参与性，即每一个人都应当能够使用、再利用、再分发这些数据。由于开放数据概念在不断演变，开放数据还应该具备“互用性”，并基于此提出了关联开放数据(Linked Open Data)的概念。互用性的核心在于开放数据间的自由整合、关联能力，是体现“开放性”实际益处的关键，为数据的再利用和增值提供了可能。

根据数据所在领域以及数据主体的不同，可以把开放数据细化成许多分支，其中，开放科学数据(Open Science Data)和开放政府数据(Open Government Data)是当前最热的两大领域。特别是开放政府数据，由于总量以及种类庞大繁多，与民众生活密切相关；而且因为大部分政府数据本身受法律的规定需要公开，所产生的效益也最受关注。如无特别说明，后文中提到的开放数据一般均指开放政府数据。



如果将开放数据运动比作一场马拉松，那么开放数据运动的前半程则完全由政府透明化在推动。事实上，最早开启开放数据运动的美国就是以2009年奥巴马签署的《开放政府指令》(Open Government Directive)为基础，朝着让政府更透明、让民众更好地监督政府运作的方向，推进开放数据的发展。在这个过程中，政府预算、政府支出、政府选举3项数据是开放政府数据计划中的重点对象，因此美国奥巴马政府在2014年进一步推动了《数据法令》(Data Act)的通过，从而加强了政府预算和支出数据的开放。

美国自然不是唯一通过开放数据实现政府透明的国家。根据开放政府伙伴(Open Government Partnership)计划的记录，目前全球共有超过60个国家加入了伙伴计划。而作为伙伴计划成员，需要承诺的便是开放政府数据，从而通过数据开放实现政府的透明化，帮助民众问责政府。

从2009年起，美国、英国、加拿大、新西兰等发达国家政府相继宣布了其公共数据开放计划。据美国网站Data.gov的统计(见图2.3)，截至2013年8月，全球有43个国家、160个地区已经上线开放数据或者开放政府信息的

相关站点。美国、英国、加拿大和法国是开放数据的先行者，且数据的可用性较高；新加坡、丹麦、意大利、新西兰等国是追随者，处于第二梯队；澳大利亚、爱沙尼亚等则处于起步阶段，处于第三梯队。

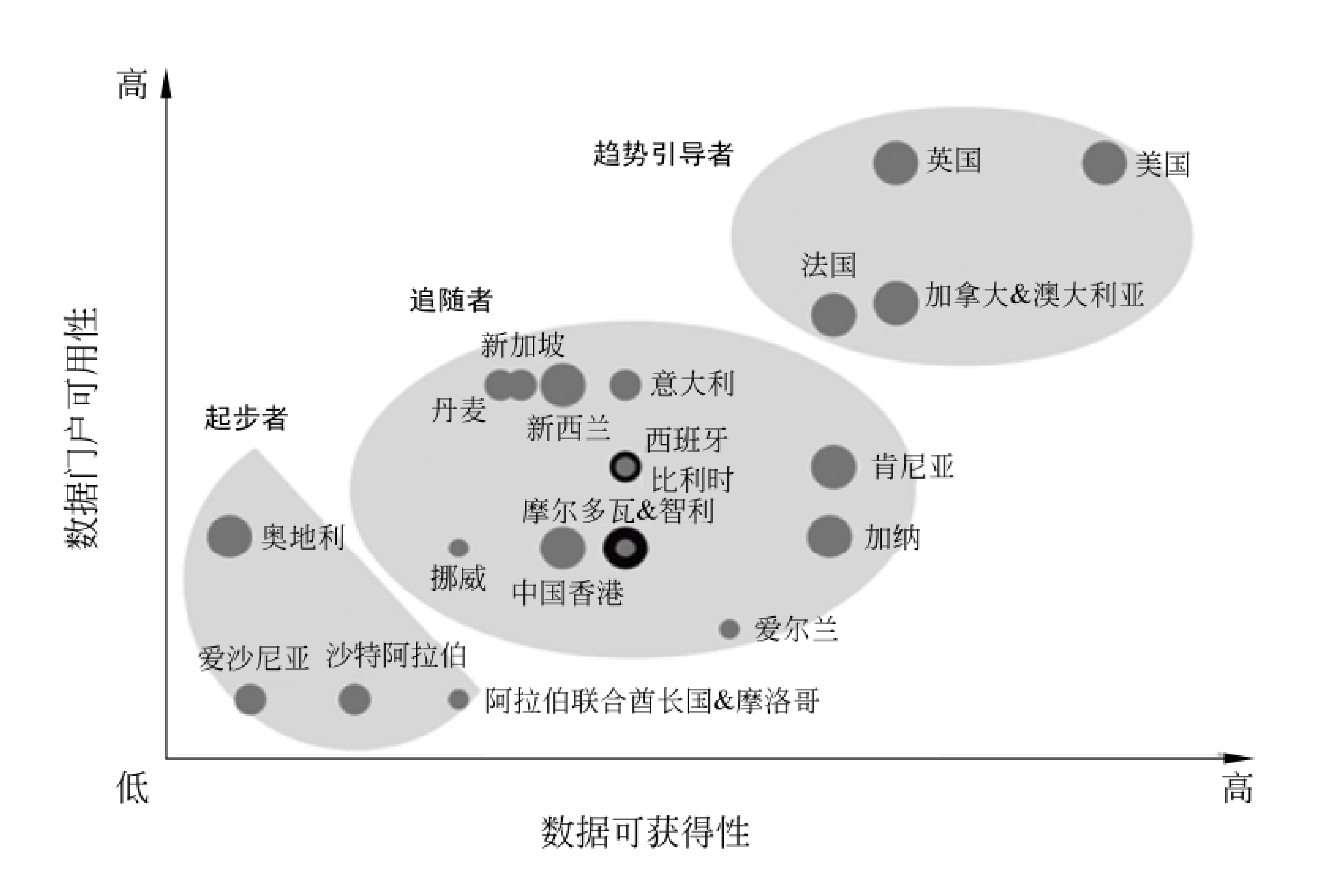


图 2.3　全球开放数据主要参与者及其角色

2013 年 6 月，八国峰会(G8 Summit)期间签署了《八国集团开放数据宪章》(G8 Open Data Charter)，简称《G8 开放数据宪章》，标志着开放政府数据已经成为全球共识。

相比国外的开放数据进程，我国的开放数据起步略显滞后。根据复旦大学国际关系与公共事务学院副教授郑磊在《中国开放政府数据平台研究：框架、现状与建议》中提供的数据，截至 2015 年 5 月，我国各地开放政府数据实践共计发布了 1963 个数据集，其中开放数据集最多的是武汉(635 个)，最少的则是贵州(17 个)。

虽然各个国家开放数据特点有所差异，但总体上可以分为三个阶段：被动开放数据阶段(1960—2009 年)、主动开放数据阶段(2009—2011 年)和挖掘数据价值阶段(2012 年至今)。

222 开放数据的社会化利用

开放数据运动产生于大数据汹涌发展的背景下并非偶然，因为政府数据的权威性、公益性和全局性，使之成为大数据发展的重要补充和落地应用手段。例如：芬兰的 tax free 项目和英国的 where does my money go 项目都向民众展示了政府如何使用税收；丹麦的 folketsting. dk 项目追踪议会动态以及立法进度，公众可以清楚地知道议会发生了什么，哪些议员参与其中；ODI 的商业计划中提到，加拿大政府靠开放数据挽救了 32 亿加元的慈善税收因诈骗造成的损失。

开放数据能够成为公民监督政府的有力工具。例如：美国加州政府就将金融危机的纾困款项公之于众，居民可以上网浏览每个地方行政单位所得到的经费。

开放数据的实施有利于提高政府部门之间协作的效率。开放数据的包容性打开了政府内各部门、政府与民众之间的边界，信息孤岛现象不再存在，数据共享成为现实。一方面推动了政府各机构开放创新，政府各部门开展业务数据分析，发现数据背后隐藏的模式和微妙关系，用新思路、新方法、新举措破解经济社会发展过程中遇到的各种问题，也成为创新的主体。另一方面政府各机构提供数据、问题和激励，邀请社会公众共同解决问题，通过众包的形式激发了大众的智慧，推动了社会创新。

以美国纽约市为例，2012 年 2 月纽约市通过了《开放数据法案》，当年 3 月由市长迈克尔·布隆伯格签署后正式生效。这是美国历史上首次将政府数据大规模开放纳入立法。之后随着详尽犯罪记录数据的开放，不仅开发出了提示公众避免进入犯罪高发区域和提高警惕的手机应用，从而降低了犯罪发生的概率；而且还能将犯罪记录信息和动态交通数据结合起来，起到指导调配警力的作用。公共交通系统的动态数据公布后，随之许多学者和商业机构分别对其进行深度挖掘，不仅创造出了手机应用，为公众出行提供实时建议，而且为地铁系统在客流高低峰时段对热点站和普通站之间的调

配提出了更优的方案。这在原来警察局或交通部门各自垄断数据的情况下是不可想象的。

开放数据一方面推动传统企业转型，另一方面也催生了许多新的中小企业。例如：丹麦的 husetsweb. dk 可以帮助用户找到提升家庭能源使用效能的方法，包括提供财政规划和联系施工承包商；英国的 Mastodon C、Carbon Culture 和 Honest Buildings 等都是利用开放数据提供服务的创新企业；Google 翻译服务使用了海量的欧盟多国语言文档来训练其翻译算法，进而提升了服务质量。

美国的 Zillow 公司作为一家市值 30 亿美元的公司。它创建了一个在线房产交易平台，供房屋产权人、购买者、售卖者、租赁者、中介、出租者、贷款经纪、房产经理等发现并分享房产及周边相关（如贷款）信息。整个平台由一个记录了超过 1.1 亿条美国房产记录的数据库驱动，这个数据库中既有挂牌出售的房产信息，也有未挂牌的房产信息。Zillow 将美国政府开放的土地交易记录、房屋交易记录、房屋整修记录、治安状况等有关社区状况的开放数据整合进原有平台，建立了更为合理的房屋估值模型。

美国的 Climate Corp 公司于 2013 年 10 月被 Monsanto 公司以 9.3 亿美元收购。它利用美国政府免费开放的 60 年农作物收成数据、美国超过 100 万个气象监测站的气象数据以及 14 TB 的土壤质量数据，为农民提供农业种植和金融决策辅助信息。它的一款主打产品是“全气候保险”，这款保险产品将在系统预测有恶劣天气时自动赔付农民的损失，而不需要农民举证实际损失。Climate Corp 公司是 2006 年由两名前 Google 公司员工创建的，它每天要利用从 22 个数据集中经过高级数据分析产出的 300 万份数据资料。这些数据来自不同的第三方机构，比如美国气象局，并且这些数据都是免费、自由重复使用的。

美国的 Mastodon C 公司成立于 2012 年 4 月，可帮助企业运行一个可定制的零碳云基础设施，并通过先进的分析建议帮助其客户释放数据的潜力。其创始人 Francine Bennett 认为，开放数据对创业企业的价值是难以估



量的。Carbon Culture 提供数字服务，可帮助企业加强沟通，实现员工参与和直接节约能源，使其转向可持续发展。

美国的 ITriage 公司是一个初创企业。它由一个急诊室医生创建，目前雇用 90 个员工。该公司的主要业务是，运用从美国卫生与人力资源服务部(HHS)下载的关于健康医疗提供者的位置和特点信息，开发一种移动应用。这种应用可以帮助 800 万民众发现满足其需求的、最适合的当地医生和医院，从而帮助人们挽救了生命。

美国的 OPower 公司的业务是运用政府发布的有关能源使用、天气和能源设备效能的数据，为客户提供节约能源的个性化建议。该公司目前雇用员工 200 多人，已帮助居民客户节约 1.4 小时能源时间(相当于一个小城市全部家庭一年的用电量)和 1.65 亿美元的能源。

223 开放数据的推进模式

1. 发布机器可读的高价值数据和推动数据的开发利用是开放数据的两大重点

开放数据要满足几个条件：一是每个人都可以获取，二是机器可读，三是不需要成本就可以获得，四是对数据再使用和分发没有限制。可见，开放数据的关键是更多的信息发现和信息利用。因此，开放数据并不是简单地将数据电子化、格式化，怎样降低获取数据的难度和提高数据的再利用程度才是核心。

从国外开放数据的发展阶段来看，当前开放数据的先行国家其工作重点可以概括为以下两个方面：一是注重以机器可读的方式，优先发布高价值的数据；二是注重数据的开发再利用，采取一些激励措施，激发企业家和创新者利用开放数据开发更多的应用，从而挖掘政府数据的潜力，积极促进经济增长和创造新的就业机会。

2. 建设统一门户，逐步开放数据集

从全球范围来看，建立统一的政府开放数据门户，集中开放可加工的

数据集是各国数据门户网站的一个普遍做法。各国数据门户网站域名中都普遍带有“数据”和“政府”字样，如 data. gov（英语）、datos. gob（西班牙语）等。在门户网站上，重点开放机器可读的数据集（datasets）、应用程序（APPS）等资源，有些数据门户网站上还设置了供开发人员参与和公众反馈的专栏。

在全国范围内建立统一的开放数据门户是各国通行做法，但是由于国家间的差异，有些地方和部门也建立了单独的数据开放门户。例如：新加坡采用的是统一数据门户网站 data. gov. sg，截至 2014 年 2 月 10 日，门户网站上开放了 68 个部门的 8733 个数据集，实现了全国范围内的整合；美国的数据门户 data. gov 在 2014 年 1 月全面改版，截至 2014 年 2 月 10 日，网站上共开放了 88 137 个数据集、349 个应用程序、140 个移动应用，参与的部门达到 175 个。

除了在国家数据门户上整合了部分州、地方政府的数据集外，美国还有 40 个州、44 个县市建立了单独的数据门户。美国的数据开放格式多达 46 种，其中应用最广的格式是 HTML、ZIP 和 XML 三种，数据集分别有 20 775 个、12 517 个和 11 992 个。英国除了全国统一的数据门户网站外，伦敦、曼彻斯特等地以及索尔福德市议会等 16 个地方和部门也建立了独立的开放数据门户。在英国的数据开放门户网站（data. gov. uk）上，共开放了 13 670 个公开的数据集以及 4170 个非公开的数据集。

各国开放的数据集以 CSV、HTML、XLS、NII、PDF 等一种或多种格式出现。在印度，目前使用的是全国统一的数据开放门户网站（data. gov. in），共开放了 5811 个数据集，共有 58 个部门和 4 个州参与，开放了 24 个应用程序；在 5811 个数据集中，以 XLS 格式开放的有 1793 个，以 ZIP 格式开放的 4 个，以 CSV 格式开放的 2087 个，以 HTML 格式开放的有 30 个，以 XML 格式开放的有 1897 个。

3. 围绕民生的地理、交通等领域最先取得成效

数据开放运动的一个核心目的就是更好地满足公众的需求，通过政府

开放数据，促进公共服务领域提供更好的服务，通过政府数据的免费使用来带动创新，创造出一些有助于大众更好地适应现代生活的实用工具和产品。

2013年6月17～18日，美国、英国、法国、德国、意大利、加拿大、日本和俄罗斯八国领导人在英国北爱尔兰厄恩湖举行了2013年八国峰会。会议围绕全球经济增长与就业、开放贸易、税收体系、土地交易、开放数据、粮食安全、气候变化、反恐问题和外交政策等展开讨论，同意在扩大自由贸易、打击跨国企业逃税避税、提高政府和商业运作透明度三个领域采取行动，并发表了联合公报和声明。会议期间，八国认可了《八国集团防止公司信息滥用和立法安排主要行动原则》，签署了《八国集团开放数据宪章》，并将上述两个文件作为附件列入联合公报。

《G8开放数据宪章》明确了开放数据的5大原则和14个重点开放领域，其主要宗旨是推动政府更好地向公众开放数据，挖掘政府拥有的公共数据的经济潜力，促进经济增长，激发创新，以及加强责任感。《G8开放数据宪章》及其技术附件的关键要点请见本书附录B。

从各国开放数据门户情况来看，围绕民生需求的数据在开放数据中比重最高，也颇受用户欢迎，但是民众关注的热点与国家的社会体制和经济发展情况密切相关。如：美国新版的数据开放门户，将原来的金融、企业、农业、海洋和安全等六大类数据集拓展至农业、消费、教育、能源、金融、地球空间、全球发展、医疗、就业和技能、公共安全、科研、气候、企业、道德、法律、制造、海洋、州、市、县等二十大类，与民生需求相关的数据集普遍增加。

在加拿大，下载量最高的十个数据集中有九个来自加拿大公民身份与移民局，包括永久居民的申请流程和时限、永久居民的分类、等待中的永久居民申请等。在新加坡，阅读量最高的数据集为人民协会总部、3G移动用户数、各运营商3G移动通信服务平均速率。在印度，下载量最高的数据集为电子和计算机科学的技术发展、印度国防研究与发展组织的热成像产品、国内储蓄及构成占GDP现价的比例等数据集。

2.3 政府信息资源开发利用

2.3.1 行业大数据的协同应用

2.3.1.1 交通领域

通过开放公共交通数据，政府将允许第三方人员使用这些数据来创建应用程序，以此改进市民出行体验；市民也能够使用开放数据报告基础设施出现的问题。许多互联网企业正在做类似的事情。例如：谷歌设计了MapMaker，任何人都可以在谷歌地图上做标注。基于这款产品形成了“在线公民制图员联盟”，花了两个月的时间，就将巴基斯坦地区长达25 000多千米的未标注公路线绘制出来了。

1. 美国交通部开放数据改善交通效率和安全性

提供大量的业务服务，致力于解决与国家运输系统相关的复杂安全问题是美国交通部的工作重点之一。美国交通部长期以来重视向公众公开数据工作，根据美国《开放政府指令》的要求，该部门先后于2010年6月和2012年4月制定发布了第一个《开放政府行动计划》和第二个《开放政府行动计划》。美国交通部已经在Data.gov门户网站上发布了765个数据集或工具及大量的应用。

美国交通部的管理者意识到，为了更好地提供高价值的数据，有一些基础性问题必须通过内部政策得到解决，包括制定数据清单，选择合适的数据进行发布；研究决定如何建立整个交通部范围内的数据架构；以对产业和个人有用的方式提供数据，同时要遵守安全、隐私和保密的相关规定；保持数据的质量，并与利益相关者形成关于数据可用性方面的对话机制等。

为积极应对上述问题，更好地完成《开放政府指令》提出的目标要求，美国交通部主要采取了两方面行动。





一是组建强有力的工作团队。2010 年,美国交通部指定负责 IT 政策监督的副首席信息官为该部负责开放政府的高级负责官员,并成立了开放政府工作组,该工作组由政策、预算、绩效、战略规划、人力资源管理、技术运营和法律等方面的专家组成,其一大任务是向高级领导者提出政策战略建议,建立美国交通部开放数据政策。

二是研究制定战略性行动计划并积极推进实施。在该部负责开放政府高级官员的带领下,2010 年 6 月研究制订的第一个开放政府行动计划成为其他机构学习的模板。该计划主要从以下三个维度推进工作。

(1) 战略维度:在短期内,要转变交通部对信息发布的态度;从长远来看,要实现和维持开放性。

(2) 政策维度:提供确认数据集和按照优先顺序发布的指导。

(3) 目录清单:建立信息系统资源列表,运用这些列表形成一个完整的数据集目录清单,并按照优先次序排出对外开放的数据。

美国交通部在 2012 年 4 月制订的第二个开放政府计划中确定了"安全社区"旗舰项目,即在 Data. gov 门户网站上开设安全社区版块(Safety . Data. gov),当时在该版块中提供了 713 个数据集、4 个移动应用、14 种资源和公共软件工具和 3 种挑战比赛。2012 年 9 月,召开了首届安全数据大型论坛(Safety Datapalooza),其目标是运用大量的安全相关的数据集,通过挖掘安全应用开发者的创新、互联网的即时性和政府收集的相关安全信息,使公众能够在大量对安全现状描述和影响将来安全环境分析的数据基础上,做出更好的与安全相关的决策,从而提高美国公共安全水平和改善公共健康。美国交通部的开放数据及相关应用如表 2.1 所示。

下面重点介绍两个应用:一是美国国家公路交通安全署的 SaferCar APP,二是航班延误时间的分析系统 Flyontime. us。

1) SaferCar APP

SaferCar APP 在美国国家公路交通安全署网站 SaferCar. gov 上向消费者提供实时汽车安全信息,主要包括:

表 2.1 美国交通部开放数据及相关应用

数 据 集	API	APP
轮胎质量(Tire Quality)	铁路设备事故报告API(Rail Equipment Accident/Incident Report API)	紧急响应指南 2012 移动 APP(Emergency Response Guidebook 2012 Mobile APP)
安全等级(Safety Ratings)	公路-铁路路口事故API(Highway-Rail Grade Crossing Accident/Incident Report API)	公交更安全 APP(SaferBus APP)
儿童安全(Child Safety)		汽车更安全 APP(SaferCar APP)
法律数据(Law Data)		道路安全展示板和社区实践(Roadway Safety Dashboard and Community of Practice)
执行和遵守数据(Enforcement & Compliance Data)		铁路安全数据可视化(Rail Safety Data Visualization)
其他数据		

(资料来源：工业和信息化部电子科学技术情报研究所)

(1) 5 星级安全汽车排名信息。考虑购买汽车的消费者可以查找事故测试排名,且可在不同品牌和车型之间进行对比。



(2) 召回信息和投诉信息。APP 用户可预见可能碰到的安全问题。如果发现安全问题,消费者可进行登记并由美国国家公路交通安全署发布公告,从而使消费者就可能存在的问题向美国国家公路交通安全署投诉更加容易。

(3) 提供安装小孩座位帮助。APP 用户可快速确定最近的小孩座位检查地点,并得到相应帮助。

(4) 安全头条和警告信息。APP 用户可从美国国家公路交通安全署获得重要新闻和信息,包括召回通知,并可推送被记录汽车通知。消费者可以利用这些数据和信息做出购买决策,确保购车安全、开车安全和维护安全(buy safe, drive safe, stay safe)。

2) Flyontime. us

Data. gov 上线以后,美国交通部开放了全美航班起飞、到达、延误的数

据，有程序员立刻利用这些数据开发了一个航班延误时间的分析系统（Flyontime.us）。该系统向全社会免费开放，任何人都可以通过它查询、分析全国各次航班的延误率及机场等候时间。这个系统上线之后，由于其简单、实用，获得了全美多个新闻报刊的报道和关注，成为很多人乘机、候机的行动指南。

以波士顿至纽约的航线为例，用户可以在系统主页上通过机场名称查看不同天气、不同日期、不同时段、不同航空公司、不同航班等各种条件下飞机是否准时以及平均延误时间的数据明细。这个简单的操作，对消费者和整个社会的经济活动具有巨大的作用。

（1）帮助消费者找到表现最佳或者最符合自己需要的航班。如果没有这些信息，消费者在选择航空公司的时候，信息是不完全、不充分的，与航空公司构成一种典型的信息不对称关系。航班的历史数据很有参考价值，公开这些信息，弥补了消费者的信息不对称。此外，消费者在对比分析大量历史数据的基础上，自己做出判断，即使结果不尽如人意，也会感觉公平。

（2）最大程度降低了旅客等待时间的不确定性。憎恶等待，是人之常情，因为等待意味着时间流失、经济损失，不确定性的等待还往往导致精神焦虑。单次航班的延误时间似乎是随机的、无规律的，但是，当数据累积到一定程度时，航班延误时间的长短就会在统计上呈现出一种秩序和稳定。航班延误分析系统把这种统计学上的“秩序和稳定”传达给了旅客，帮助他们建立正确的期待，合理安排时间，避免焦虑。

（3）有利于推动航空市场的良性竞争。航班延误分析系统按平均延误时间给相关航空公司排了“座次”。回到上面的例子，经营波士顿至纽约航线的公司共有5个。就是否准点而言，谁好谁差，几乎一目了然。此外，各次航班的表现也有明细。例如American Eagle航空公司的第4617航班，全年共有182班次，平均延误7分钟；相比之下，该公司的4614航班，全年也是182班次，但平均提前8分钟到达。这些数据，不仅是消费者的行动指南，也是各大航空公司的核心竞争指标。通过公开这种数据，无疑可以促进市场

竞争，航班延误必然逐渐下降到消费者能够接受的合理范围之内。

其实，为了缓解航班延误的问题，美国政府也想过同样的办法。早在 Data. gov 之前，国家交通安全局就在其网页(NSA. gov)上提供过一个“航班等待时间计算器”，帮助旅客估计因航班延误而导致的等待时间。2009 年政府开放数据之后，民间开发出来的这一免费工具明显比交通安全局提供的“计算器”功能更强大、界面更友好。很快，该局便关闭了这个“计算器”，也节省了维护这一应用的开支。

Flyontime. us 还能够查询各个机场安检通关的时间，这个数据也是机场服务质量的一个重要指标。但这部分数据来源并不是政府发布的数据，而是乘客自己提交的数据。候机的乘客可以通过推特(Twitter)或者智能手机向该系统提交其在某个机场通过安全检查的时间。这些数据，通过汇总和平均，成为其他用户的参考。

伴随着 Data. gov 的开放，美国的航班延误率正在呈下降趋势，由 2008 年的 27%下降到 2009 年的 20.8%，再到 2010 年的 20.2%。数据开放在其中的作用不可小视。

2. 旧金山利用开放数据优化城市交通系统

美国旧金山市为公众提供了大量的开放数据，包括从停车计费器到公共艺术表演等信息，但并不是所有的人都意识到了这些信息的作用。为了让民众进一步了解这些开放数据的价值，旧金山市联合 Kicker Studios 公司，通过对开放数据的利用来优化城市的交通系统，见图 2.4。

Kicker 公司拥有大量能够用来处理的公开数据，包括公车路线、事故报告、最快和最安全的路线以及停车信息等。在与旧金山运输局进行会谈之后，Kicker 公司发现，公车信息更新系统 NextBus 中所运行的日程方面的数据每年只更新四次，这就意味着每次公车的延迟情况(这种情况其实很常见)都可能会造成车次抵达时间的偏差，而运输局却没有中央通信系统，司机们只是在工作结束之后把这些延误情况记录在事故报告中就算完事了。

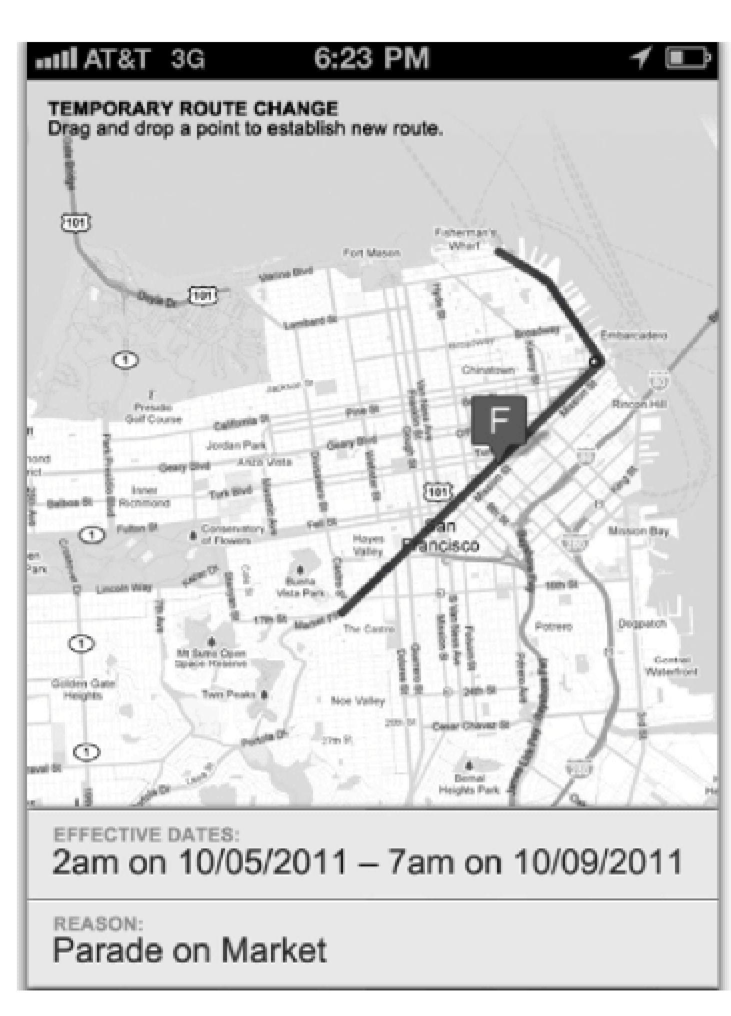




图 2.4 Kicker 利用旧金山开放数据开发的应用

因此，Kicker 公司建议使用一个短信息的界面来记录事故报告，同时让司机们来重新设定自己的公车路线。也就是说，当一场球赛结束后，民众对公车的需求会比较强，但如果按照原有的路线是不能为当时所有等车的人提供运输服务的。而针对陈旧的公车调度系统，Kicker 公司也提出了一个非常全面的解决方案，界面上包括实时的交通信息和正常的公车路线，可以在公车进站前对其到达时间进行更好的预估。同时，该界面还允许用户丰富公交车的相关信息，比如在 14 号线上有一位喜欢吵架的乘客之类的信息。



图 2.5 Kicker 公司开发的应用用户界面

图 2.5 中显示的是 Kicker 公司开发的应用中一个基于手机或网络的用户界面，它能够展示距离自己最近的公交线路、下一班车到来的时间、出租车最多的线路(有时候出租车是很难找到的)和最佳的骑车路线等。此外，该界面还能够告知用户最

省钱和最省时的交通方式。所有这些服务都需要用到开放数据。

3. 移动应用SpotHero缓解多个城市停车难问题

SpotHero是一个手机应用，支持iOS和Android手机，能够实时跟踪入网城市的停车位数量变化，用户只需要输入地址或者在地图中选定地点，就能看到附近可用的车库或停车位以及价格和时间区间。目前它已经能够实时监控包括华盛顿、纽约、芝加哥、巴尔的摩、波士顿、密尔沃基和纽瓦克七个城市的停车位。

4. 芝加哥市推出"领养"人行道的应用

芝加哥市推出了"领养"人行道的应用，市民志愿者将在大雪天为自己领养的人行道清除积雪，不仅方便居民出行，而且减轻了市政开支。

5. 里昂市用交通数据治堵

2013年IBM的研究者与法国里昂市合作开发了能缓解道路拥堵的决策支持系统优化器（Decision Support System Optimizer，DSSO），基于实时交通报告来侦测和预测拥堵。若交管人员发现某地即将发生交通拥堵，就可以及时调整信号灯让车流以最高效率运行。这个系统对于突发事件也很有用，例如帮助救护车尽快到达医院。随着运行时间的积累，这套系统还能够"学习"过去的成功处置方案，并运用到未来预测中。

6. 浙江某市利用大数据改善交通管理

浙江省某市经济发展迅猛，地方交通越来越繁忙，机动车辆不断增加，经过几年交通信息化的发展，已经陆续接入了100多套智能监控卡口系统、300多套卡口式电子警察及500余路视频监控。地方交通部门采用了数据驱动的方法，在市内重要检查点安装了上千台数字监控设备，这些设备每周7×24小时不间断地捕获图像和视频数据，每月数据量达TB级。这些数据采集设备获取的结构化数据，例如：时间、地点、车辆信息等集中存储在市交通支队数据中心，而图片和视频等半结构化数据存储在各县数据中心。当地交通部门面临着如何有效利用这些不断增加的交通信息数据改进交通管

理的挑战。

(1) 集中管理交通数据：集中访问分散存储在不同的支队数据中心的图像或视频等交通数据和道路交通管理设施、装备及应用系统等。

(2) 优化海量数据利用：提供尽可能长时间段的车辆监控数据为市公安治安、刑侦、经侦部门人员及一线民警等提供信息支撑服务。

(3) 改善交通：提高对各种交通突发事件的应急调度能力，依据历史数据预测交通或突发事件的趋势。

其解决方案主要包括三个方面：

(1) 部署统一的交通数据中心：通过22台服务器、198TB的存储空间对数字交通信息实现集中存储。

(2) 部署Apache Hadoop软件：利用Hadoop分布式文件系统(Hadoop Distributed File System，HDFS)和Apache HBase实现基础过车结构化数据的永久存储以及最近24个月的交通违法图像数据，实时检索数据，并可随时无缝扩容。



(3) 部署城道重点车辆动态监管系统：发挥开放数据分析平台的优势，实现海量数据的挖掘和分析。

以上方案取得了显著的效果：

(1) 提升了交通案件侦破能力。机动车违法图像信息在系统的保存周期从3个月延长到24个月，交通警察等部门可根据车辆的颜色、车型、号牌等信息实时查询其历史行为、行车路线和车辆营运公司、驾驶人等关联信息。

(2) 增强了交通警察对机动车辆的监管能力。交警可以从24亿条过车数据中轻松检索被监测机动车的号牌，精确查询行车轨迹。

(3) 便捷利用关联车辆的分析数据。针对24亿条实际过车数据进行两卡点、多卡点的伴随车辆和碰撞车辆的复杂分析，查询耗时仅为10秒左右。

2.3.1.2 医疗领域

开放医疗数据可以帮助实现类似大规模流行病分析这样的研究，并产生实质性的突破成果。在这个过程中，需要严格制定措施，确保病人的隐私权利。例如，可以向研究人员开放出生时的健康状况登记，允许医生通过追踪丰富的信息，发现环境因素对人体健康的影响。谷歌流感趋势(Google Flu Trends)数据已经展示了通信连接和规模化两者结合可以改变我们对某种已知病毒的了解，几乎不用再分享和核对信息。

简单地集中管理数据并将其向研究人员和临床医生开放，就已足够医疗人员开发出可更好地了解和治疗疾病的新模型，医疗保健机构通过将病人的健康记录电子化及隐私化，同时为医生、保险公司、相关部门和病人开放数据，可以产生更大的价值。各种数据与电子病历记录相连接又可挖掘出新数据，包括病人满意度调查、医生的临床笔记以及磁共振的图像数据等。健身和健康追踪器产生的大量数据看似有趣，但个人很难从中搜集到有意义的东西；然而，当成千上万人的数据被用于挖掘与健康结果相关的信号和链接时，这些数据就可能发挥作用，比如可被用作预防疾病或及早检测到疾病的新方法。

美国食品及药品管理局(FDA)正在为某些药物发行标签，为经历某种基因变异的病人注明不同摄入剂量(或解释为何某些病人不能服用该药物)。这预示着未来可能实行更多个性化的用药措施。医院开始使用的Clipmerge软件更便于医生进行快速查找，同时，当电子医疗记录表单上的药物可能发生相互作用时，也能及时通知医生。

1. 美国卫生和公众服务部

美国卫生和公众服务部是美国政府最大的卫生保障机构，是美国医疗系统的官方最高管理机构，基本职能是保护国民身体健康，提供最基本的医疗卫生服务。该部非常重视开放数据工作，是第一批发布机读版数据目录的联邦政府行政部门之一。该部于2010年6月发布了第一个《开放政府计

划》,从领导治理与文化变革、透明、参与和协作、旗舰行动4个方面进行了详细计划,提出了5大旗舰项目和80多项专门工作,并且在hhs.gov/open网站上周期性地发布实施进展情况,如表2.2所示。2012年4月,该部又正式发布了第二个《开放政府计划》(计划于2013—2014年完成),在第一阶段工作基础上又明确了3大旗舰项目和60多项专门工作,并增加了“智能发布”和“大数据”的两个专门项目。截至2012年4月,美国卫生和公众服务部已经在Data.gov网站上提供了300多个数据集和工具。在第一个开放政府计划中提出的里程碑要求都已实现或超过,正在以新的方式带给公众更多的收益。美国卫生和公众服务部于2009年成立了创新委员会,主要负责协调和监督该部的《开放政府计划》,推进该部朝着更加透明、参与和协作的方向努力。美国卫生和公众服务部还成立了一个包括创新委员会成员的工作组,定期评估信息获取方面的工作。该部在开放数据工作中注重公众参与和协作,自从发布第一个《开放政府计划》以来,已经组织了50多个挑战性竞赛,以吸引公众参与到问题讨论和解决方案的建设中来。



表2.2　美国卫生和公众服务部《开放政府计划》中的旗舰项目

(a) HHS《开放政府计划1.0》中的旗舰项目

1	医疗保险和医疗补助展示中心(Centers for Medicare & Medicaid Dashboards)
2	食品和药品管理局透明度行动计划(FDA Transparency Initiative)
3	食品和药品管理局透明度、结果、责任、信誉、知识共享计划FDA-TRACK (Transparency-Results-Accountability-Credibility-Knowledge-sharing)
4	蓝色按钮行动计划(“Blue Button” Initiative)
5	社区健康卫生数据行动计划(Community Health Data Initiative)

(b) HHS《开放政府计划2.0》中的旗舰项目

1	创新团队项目(Innovation Fellows Program)
2	旨在提高医疗创新的新合作项目(New Collaborations to Promote Medical Products Innovation)
3	提高数据质量和可用性(Enhanced Data Quality and Usability)

(资料来源:工业和信息化部电子科学技术情报研究所)

下面重点介绍两个重要项目:一是健康指数仓库,二是健康卫生数据行

动计划。

1）健康指数仓库（Health Indicators Warehouse）

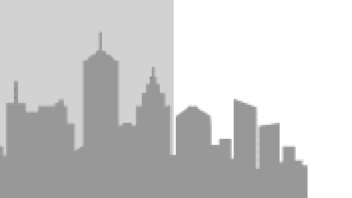
数据指数仓库旨在提供一个有关国家、州和社区健康指数的单一、界面友好的数据来源，满足多人口健康行动的需要，作为数据港为健康公共卫生数据行动计划提供服务。数据指数仓库项目由美国健康统计中心（美国卫生统计中心）开发维护，提供数据支持和资金协作的部门有美国医疗保健和医疗补助服务中心、美国卫生和公众服务部副部长办公室、青少年健康办公室、疾病预防和健康促进办公室、少数民族健康办公室、计划和评估助理部长办公室。可从该项目网站上分主题（如疾病、条件、年龄等）、地理（州、郡、医院等）和行动角度获取相关数据。2013 年 6 月 1 日，该网站发布了最新版（1.7 版），提供 1215 个指数。

2）健康卫生数据行动计划（Health Data Initiative）

该项目原名为社区健康卫生数据行动计划（Community Health Data Initiative），后改名为健康卫生数据行动计划（Health Data Initiative，HDI），是美国卫生和公众服务部开放政府中的一个主要的成功项目，其核心工作是提供中央数据资源，以帮助新的数据用户确认可以创造新应用和服务的数据。HDI 由美国医学研究所、美国卫生和公众服务部在一个会议后于 2010 年联合发起，该会议的参会代表包括来自联邦政府部门、学术机构、社会领域、公共卫生社区、信息技术公司、主要业务企业和保健实施系统的领导。目前，健康卫生数据行动计划已组建了联盟，现有 17 个单位会员，并设有联盟网站。该项目的目标不仅要有效配置数据，而且要引发创新和对那些新应用的使用，该项目通过公私合作，鼓励创新者应用健康卫生数据开发各种应用，提高健康卫生意识和改善健康卫生成效，激励改善健康卫生的社区行动，帮助美国民众更好地理解健康卫生和所在社区的健康卫生保健实施情况。该项目开展的工作有：美国卫生和公众服务部发布大量更加可用的健康卫生数据，软件开发者使用健康卫生数据开发新的应用，随着数据的不断改善和新应用的不断开发，消费者、社区和服务提供者在疾病预防、健



康卫生促进、保健质量提高等方面取得了新的成效。该项目已扩展到许多州和地方政府的社区。美国卫生和公众服务部门在倡导建立数据使用者和提供者这一生态系统方面发挥了重要的领导作用，这个生态系统为改善政策制定者、公众、健康医疗专家、研究者和其他人员的决策创造了价值。Healthdata.gov 网站为创新者生态系统提供一站式资源，这些创新者将数据转换成新的应用、服务和观点，帮助人们改善健康。该网站的用户能够免费获得与健康相关的数据，查找与健康相关的创新，并可与其他创新者联系，通过网站上的专门栏目咨询有关数据集的专门问题，通过应用程序接口获取所有的数据集目录。



健康卫生数据行动计划每年举行一次名为“Health Data Palooza”的大型医疗卫生数据行动年度论坛，为数据使用者、数据拥有者、开发者、风险投资、政府和企业提供交流如何挖掘数据潜力的机会。2013 年 6 月 3～4 日在华盛顿举办的第三届论坛内容丰富，包括产业界和政府官员的重要讲话，数据使用的专题讨论，50 多家单位的新应用展示（App Demo），由开发者、诊所和技术专家参加的额度为 2.5 万美元的编码比赛（Health Code-a-Palooza），以及发布联盟下一步将开展的重要行动计划和项目，宣布挑战比赛的获胜者和即将发布的数据集等。

健康卫生数据行动计划正在采取的措施主要包括：

（1）向公众免费提供来自联邦、州、地区和郡的有关保健、卫生和医疗成效的数据，这些数据没有知识产权限制、容易获取，具有标准化、结构化的特点，其中有年龄、性别、种族、收入的数据，也有来自医疗保险和医疗补助服务中心的关于疾病、质量、费用等的数据，以及以前从未发布过的数据。

（2）与技术公司、研究人员、卫生提倡者、媒体、消费倡导者、市场人员等进行广泛的沟通交流，帮助激励民间领袖和公众改善公共卫生的状况。

具有潜力的案例有：

（1）交互式医疗卫生地图，使公众能够方便、清晰地了解其所在地区与其他地区的医疗卫生情况的比较。

（2）“排行榜”使市长和其他民间领袖能够跟踪和了解当地医疗卫生现状和存在的问题。

（3）社交网络应用能使医疗卫生领导者与其他人员互相联系，比较成效，共享实践经验。

（4）在线游戏能够帮助教育民众了解社区医疗卫生情况。

（5）PatientsLikeMe（像我一样的患者）是一个社交网络医疗站，建立在美国医疗服务部（US Department of Health Services）的开放数据之上，可让潜在患者有机会及早发现病情，也可让更多病人互相分享信息，彼此交流病症情况。

2. 西奈山医疗中心

西奈山医疗中心是美国历史最悠久、规模最大的教学型医院之一，其在医学教育与生物医学研究方面的地位非常突出。目前该中心正利用来自大数据新兴企业Ayasdi公司的技术对整个大肠杆菌基因组序列进行分析，其中包括超过100万个DNA变异，旨在努力理解某些菌株如何在与抗生素的共处中获得抗药性。细菌的抗药性影响着全球各地数以百万计的病人。Ayasdi的技术为数学研究、拓扑数据分析（简称TDA）开辟了一片新天地，有助于人们更深刻地理解数据形态。西奈山医疗中心的目标是用这些方法为多种疾病的病人进行个性化诊断和治疗，比如癌症和糖尿病病人等，以及改善医院对病人的护理情况。

在预测方面，西奈山医疗中心已经将名为PACT的预测模型内置于电子医疗记录系统，用以预测出院病人90天内返回医院的可能性（新卫生保健法为医院提供了某些财政奖励，以减少90天内病人的再入院率）。根据预测，医疗中心的高风险病人或许将真正获得不同的护理，比如为他们分配一个治疗后协调员。

3. 数据开放造就医生图谱

在ZocDoc、Healthgrades、Vitals、Yelp中虽然可以找到医生在病人中的

口碑，但是，病人对医生的评价毕竟还是会存在一定的片面性和主观性。如果一位医生在其他医生中的口碑也不错的话，那这位医生应该就错不了。

自称为“黑客活动家”的 Fred Trotter 通过 FOIA（信息自由法案）申请到了数百万份美国联邦医疗保险（Medicare）的医生推荐数据，然后将这些原始数据提供给 Medstartr 大众融资活动的支持者，成功募集到 1.5 亿美元。他还发动了将当前数据集与另一个数据集进行合并的活动，以打造“医生图谱”。

2012 年 11 月，医患网络初创企业 HealthTap 发布了一项名为 DOConnect 的新功能，该功能将 HealthTap 自身的医生数据（约 17 000 名医生）与 Trotter 拿到的联邦医疗保险医生推荐数据等结合起来，为病人展现出一个全新视角的医生互信网络。

此项功能可以让病人了解到 250 万名医生的推荐关系，每一名医生的关系和位置可尽收眼底。这些信息可以帮助病人在搜索医生和专家时做出决定，同时也可以让医生有机会建立一个反映其线下网络的在线网络。



Trotter 的目标是创建一套病人感觉有用、医生认为公平的排名算法，同时也希望学术机构、医疗政策专家、创业者能够利用这套东西来提高医疗保健的整体透明度。

2.3.1.3 教育领域

目前全世界的教师和大学机构正在以免费开放版权的形式提供高质量的教育内容。更重要的是，偏远地区的人们使用这些内容也越来越方便了，带宽和通信连接打破了社会体系中一直存在的教育壁垒。

1. 意大利教育部

意大利教育部、大学以及研究机构针对 CC-BY 协议成立了自己的开放数据门户网站，公开了意大利的学校（如地址、电话号码、网站、行政代码）、学生（如人数、性别、表现等）和教师（如数量、性别、退休等）等相关内容，旨

在将所有的数据开放使其得到更透明公开的再利用。此举将有助于意大利学校教育系统更好地被公众认可，同时为学生、教师和家庭提供更好的服务。

2.“全球教育伙伴”开放数据以评估教育进步

“全球教育伙伴”(Global Partnership for Education，GPE)组织开始于2002年的“全民教育-快速跟踪计划”(Education for All-Fast Track Initiative，EFA FTI)，致力于让所有儿童都走进学校接受优质教育的多边合作。在过去十几年中，GPE已经从7个成员国发展到接近60个成员国，调动了35亿美元扩大教育机会与提高教育质量，已帮助2300万儿童第一时间重返校园，同时支持了37000间教室的新建并培训了41.3万教师。其成员国68%的女童现在都完成了小学教育，其中18个成员国实现了入学机会的性别平等。

2013年5月，GPE宣布启动“开放数据计划”，第一批29个成员国数据在线免费开放，第二批25个国家的数据在2013年年底开放。开放的所有数据都是从淹没在GPE发展中国家成员伙伴的教育部门计划、相关部门总结文件、GPE贷款申请，以及由GEP合作伙伴，如联合国教科文组织与世界银行提供的数据中总结提炼出来，所有开放的数据都有原始来源，在注解中还具体说明了其背景和界定以及获取数据的方法。数据分6个教育大类共57项指标。六大类包括：关键教育产出与目标；国内、外部与GPE资助；学习结果，尤其是阅读与数学评估；地方教育团体的构成与发展伙伴；教育部门援助的效率。

作为监测与评价战略的一部分，GPE开发这一在线数据库的目的是对每一个GPE成员国的教育目标与实际结果进行比较，显示这些国家政府在让更多儿童走进学校、提高女童入学机会以及增加教师数量方面取得的进步，同时帮助这些国家评价其进步情况，并制定基于证据的计划解决儿童教育问题。

GPE 认为，数据的免费使用将带来巨大变化，并有助于提高成绩并影响决策制定，还能有效刺激成员国增强该国的统计系统。

3. 美国教育数据计划

美国政府于 2012 年 6 月初启动了“教育数据计划”，旨在帮助学生及其家庭从基于开放数据的创新中获益。2012 年 7 月上旬，美国白宫、美国教育部和乔治华盛顿大学商学院联合召开了“教育数据困境”(Education Data Jam)研讨会，各界教育技术专家和企业汇聚一堂，就如何利用开放教育数据开发新的应用、产品、服务及产品功能，促进学生成才展开讨论，借助“我的数据计划”(My Data Initiative)鼓励拥有学生数据的学校、软件厂商和其他机构将这些数据以电子、机器可读的格式提供给家长和学生，使学生能随时掌握自己的学习概况，获取个性化学习体验，方便他们更理智地选择学校和财政资助。

4. MOOC 教育模式

全球不断兴起的大规模开放式在线课程(Massive Open Online Courses，MOOC)教育模式，也是开放数据革新教育行业的另一例证。例如：Khan Academy 在线网站有超过 3000 份教学视频，涵盖各个方面，从物理课程到金融学指南等不一而足。全世界的人都可以使用这个不断增长的资源库，或者贡献自己的内容。通过这一平台，孟买的学生可以通过手机获得 MIT 最顶级的课程，甚至成为一名教师，上传自己的教学视频。

2.3.2 城市大数据的协同应用

在政府信息资源开发利用方面，世界上许多城市已经率先开始行动。“开源城市”已经不仅仅是互联网时代对知情权的迫切需要，它已成为政府治理方式的革新手段。

1. 基于媒体报道的 China AidData 项目

中国对非洲援助一直以来都是个饱受争议的问题，而在这长达 60 年的

资助历史中，中国对非洲到底援助了多少金额，可能没有多少人搞清楚过。致力于国际援助款透明化与开放化的研究机构 AidData 在 2012 年起便针对这个问题展开了研究。

由于中国官方并未采用一些国际援助款登记平台，如经济合作与发展组织（Organization for Economic Cooperation and Development，OECD）的 CRS 或者国际援助款透明计划（IATI）的援助款数据平台，因此 AidData 无法通过搜索这些现有的数据库来获取所有援助款数据。AidData 在这个项目中便采用了一种基于媒体报道的数据采集方式：通过对不同中外媒体源例如维基解密（WikiLeaks）、BBC 的相关报道进行挖掘与整理，AidData 研究员成功搜集了大量官方公开或未公开的对非援助项目以及金额数据，从而形成了 China. AidData 的数据库。2000—2011 年中国对非年度援助金额分布如图 2.6 所示。

图 2.6　2000—2011 年中国对非年度援助金额分布

AidData 在这个项目中证明了基于媒体报道的数据采集方式是一个很好的解密非公开援助款项目的手段。例如，非洲马拉维作为中国的受援助对象在其官方系统中仅公布了两项中国援助计划，但通过挖掘不同媒体报道，AidData 成功地将额外 14 个总值 163 万美金的中国援助项目公开，进一步将中国对非援助计划透明化。

AidData 的数据库目前囊括了 2000—2011 年间中国资助 50 个非洲国

家的 1673 个项目，项目总值达 750 亿美元。为了便于记者、研究人员、政策制定人员等访问这些数据，AidData 创建了项目网站 china. aiddata. org 提供数据的查询、访问、下载以及可视化。同时，为了提升数据的质量以及持续追踪中国对非援助项目，网站也接受用户提交的新信息，例如照片、文件、媒体报道、视频等。

2. 开源芝加哥：把整个城市搬上 GitHub

GitHub 是一个代码托管网站，但与过往许多代码托管网站不同的地方在于，其提供了充分“开放”的工作模式。它鼓励任何人对一个公开的代码库进行“复制”从而对原有代码进行修改、扩展、改正，同时，它也充分鼓励任何人参与项目的讨论，可以新开一个“工单”来提出问题，汇报 Bug，建议新增功能。正是这样“开放”的模式使其成为程序员界最重要的工具和社区。



在 2013 年 2 月，芝加哥市政府决定将其整个城市的数据上传至 GitHub，并鼓励所有人来“复制”它们的数据，帮助它们提升数据的质量或者利用这些数据做出创新的应用。这是自 2009 年奥巴马政府宣布全国开展开放数据运动及英国成立开放数据研究所以来，开放数据领域的又一模式创新。

如果说将数据放在开放门户提供民众下载是开放数据 1. 0，那么将数据放在 GitHub 这样一个鼓励开放协作的平台就是进入了开放数据 2. 0。开放协作使得数据能够像代码一样被“复制”并由社区来提升质量，而这就提供了一个“发布者-使用者”之间的双向通道来进一步帮助城市管理者将数据化为真正有用的资源，这是仅将数据开放下载所不能达到的效果。

3. 开放的城市服务热线：从 FixMyStreet 到 Open311

FixMyStreet 是英国民间非营利机构 MySociety 推出的第一款产品，也是首款在城市服务领域内引入开放模型的应用。往常，对于公共设施比如路面、街道路灯等的报修以及其他城市服务的投诉都是单向、单人的沟通，这也就造成了问题的重复投诉率高、处理进度不透明等问题。而

FixMyStreet 首次引入了开放模型，将单向、单人的沟通改造成双向、多人的沟通模式，允许多人集中对一个问题进行投诉，并提供平台对有关部门的处理进度进行追踪。

例如英国南安普敦市市民向市政府投诉有路障倒地阻碍了人行道，地图上标记了准确的问题地点，次日早上市府便立刻回复说该问题已登记在案，并且在问题解决后，立刻再次回复让公众知情。

这样的开放模型在解决城市服务问题中有着众多的优点。首先，这样的开放模型更容易吸引人参与到城市问题的投诉中。对于如今的手机党、微博党、微信党而言，简单地在地图上点点，写上两句话，要比一本正经地拨打热线电话更容易。其次，沟通成本会更低。传统的热线电话方式，使得单一问题的投诉重复率大大增加，而开放模式则使得单一问题能够由多人同时参与，这也就减轻了相关部门在接受问题投诉上所付出的时间和人力成本，避免资源浪费在同一问题上。最后，采用开放模型是政府树立良好形象的极佳途径。开放模型不仅是将工作流程开放，允许更多民众参与，更是对信息的透明化：政府何时受理该问题，是否持续跟进，是否已解决问题等信息都通过一个透明化的渠道让公众知情，而这也能更好地塑造一个透明、公开的政府形象。

FixMyStreet 的成功，引爆了一场民间对城市服务热线改造的风潮。各种类型的类似产品在各个国家、城市相继推出，民众的参与热情一度高涨，但随之产生的问题也越来越多。首先，民间自行开发的类似产品虽然可以吸引民众参与，但是有时候却无法保证政府的参与。其次，由于产品过多，政府不可能在所有产品上同时跟进问题，这反而降低了政府效率。最后，因为每个人采用的产品很可能不同，因此投诉的重复率问题又回来了，因为民众的注意力被不同产品分散了。

为了解决这些问题，Open311 诞生了。Open311 本身并不是一个新的 App，而是一个供第三方应用与政府的城市服务热线进行数据交换的 API 标准。它所制定的标准确保了各个地方政府采用统一的接口来供第三方产

品使用,这样就确保了所有第三方应用都能通过统一的渠道将数据反馈到政府机构。同时第三方应用之间也就有了统一的接口来交换以及同步数据,从而解决了上文提到的由于产品过多,民众的注意力被分散的问题。

更为重要的是,Open311 制定的 API 标准使得城市服务热线的数据得以真正开放。而此类数据对于城市规划等问题是极为重要的。2010 年,Wired 就曾经从纽约的 NYC 311 服务里私下获取过近百万 311 电话的数据,并就此制作了可视化图表进行数据分析。而现在有了 Open311 协议,通过开放的渠道来完整取得相关的数据就不再是问题了。

Open311 脱胎于美国城市服务热线 311,但它本身不仅是一个美国的标准,而是期望成为一个国际标准。目前除了美国的城市比如纽约、芝加哥之外,还有英国南安普敦、巴尼特,芬兰赫尔辛基等城市采用了 Open311 的 API 标准。

4. 用众包的 LocalData 和 Streetmix 设计城市

城市规划听上去好像是一件离老百姓很远的事情,但如果政府采用开放模型来重新组织城市规划活动,那么普通民众也能参与其中,并且还能出其不意地帮助城市规划部门提升效率。如城市规划的前期调研,规划机构往往需要耗费大量人力成本和时间成本来收集详细的城区地块数据。而这一过程如果能够让熟悉这一地块的民众来协助,则会事半功倍。2012 年,美国 Code for America 的一批成员(Fellow)在和底特律市合作过程中,便意识到了这个城市规划中收集数据的难题,进而开发了一款新的应用 LocalData。LocalData 引入开放模型的理念,由规划部门来设定详细的问题,而民众则可以通过实地考察,然后在手机应用上录入数据回答问题。

这种众包的思路在不同的美国城市都取得了极为难得的成绩。例如,印第安纳州的格雷市从 20 世纪 60 年代起就面临着人口衰减的问题,如今整座城市到处都是空宅无人居住,市政府有意将一些空宅拆除另做开发,但又缺乏翔实的数据来确定需要拆除的建筑范围,于是 LocalData 便成为解决这

一问题的关键。通过市政府和芝加哥大学公共政策学院的合作与协调，当地 67 名志愿者调查了市内 2000 英亩的 11 651 幢房屋。而调研的结果通过 LocalData 的可视化功能直观地展现给决策者，让他们了解空房的密度、空间分布情况等，极大地方便了拆除计划的制订。

普通民众的参与方式当然不仅仅局限于做这些数据的收集工作。Streetmix 是另一个由 Code for America 的成员制作的应用，旨在释放人们对自己城市街道的想象力，用简单的网页应用，通过拖曳页面元素，设计出自己心目中的街道。而人们对某一街道的设计，又能通过该平台汇总产生统计数据，例如“70%的设计中包含了一条自行车道”，那么决策者便能更好地决定是否要在新街道规划中预留出一条自行车道以及具体如何设计它。这一应用一经推出，便受到了民众的广泛欢迎。因为它本身简单易用，很多人便把它作为简易版的“模拟城市”游戏来尽情发挥想象力。例如，一位网民制作的 Streetmix 街道图就展现了把整条街都占领当作自行车道的“霸气”设想。

5. 加利福尼亚州 ISO 优化电网运行

加州独立系统运营商（简称 ISO）管理着全加州地区超过八成电网中的供电走向，每年提供的电力达到 2.89 亿千万时，惠及 3500 万民众，供电线路的总长度超过 25000 英里。他们利用 Space-Time Insight 公司的软件实现情景智能化机制，从而将多个来源的大规模数据进行关联与分析，其中包括天气状况、传感器数据以及计量设备测绘结果等，并以可视化形式帮助用户查看及理解如何对可再生能源进行优化，实现整个电网的电力供需平衡以便快速应对潜在危机。

6. 拉斯维加斯市构建实时公共事业网络模型

由于记录太过古老、信息不够准确，大部分城市中的公共事业机构都不了解埋在地下的资产处于何种状况——因此居民往往会由于某条供电线被意外切断或者某条供水管线老化爆裂而受到影响。为了解决这些难题，拉

斯维加斯市采取智能数据方式开发出一套实时公共事业网络模型。VTN咨询公司帮助市政当局通过各种渠道汇总数据，并利用 Autodesk 技术创建出实时 3D 模型。这套模型中包含着地上与地下的所有公共设施，目前已经被用于监测城市地下设施的具体位置以及运转状况。

7. 迈阿密市属戴德县使用情报仪表板节省开支

佛罗里达州迈阿密市属戴德县希望将 35 个区域自治单位与迈阿密市聚拢起来，努力帮助政府领导做出更为明智的管理决策——包括充分利用水资源，减少交通拥堵以及改善公众安全等。IBM 通过云计算环境下的深层分析为该县带来一套情报仪表板，从而帮助各机关与部门彼此协作并实现可视化管理。举例来说，戴德县公园部门 2013 年预计将通过识别并修复因锈蚀而漏水的浇灌管道节省 100 万美元经费。

8. 西雅图市使用公有云、大数据实现节能措施

西雅图市最近与微软和埃森哲试点大数据节能项目。该项目基于微软的 Azure 云计算平台，可收集和分析来自四个城区建筑管理系统的数百个数据集。通过预测分析工具，大数据系统将能找出可行的节能措施，目标是将耗电量降低 25%。

9. 波士顿鼓励个人“领养”消防栓减轻市政负担

波士顿的冬天积雪很厚，2013 年 1 月份波士顿新城区办公室发布了一款名为“领养消防栓”的应用。根据哈佛商学院博客，该项目在地图上标注了全市 13 000 个消防栓的位置，市民可以申请“领养”一个或多个消防栓，并承诺在大雪天负责将自己领养消防栓从积雪中挖出来。完成“领养”手续后，志愿者将在消防栓被雪埋时收到消息通知。

中篇 协同体系：

多源信息协同的标准、模式与架构

智慧城市是一个**开放的复杂巨系统**，城市系统本身与系统周围的环境有物质、能量和信息的交换，同时城市系统下又包含数量庞大、种类繁多的子系统。**信息协同**是保障城市系统中其他资源要素优化配置的基础，也是城市系统更加智慧运行的前提。

第3章　群体级和区域级的多源信息协同

以个体为单位，多源信息协同是实现群体一致性行动的基础。以群体为单位，多源信息协同则是实现组织联动的基础。本章分别以群体决策和城市管理为例，对群体级的多源信息协同和区域级的多源信息协同进行介绍；在此基础上，介绍智慧城市多源信息协同体系的理论框架。

智慧城市是一个不断演进的过程，通过多源信息的协同将数据的采集与业务应用分离，在决策单元分散化的基础上破解信息碎片化的难题，逐步实现物体的智能化、流程的智能化和服务的智能化，最终实现从“控制”到“智能”的转变。

3.1　面向群体决策的多源信息协同

在现实生活中，决策往往是群体行为，是由多人参加进行行动方案选择的活动，如各种委员会、董事会、代表大会等均属于群体决策机构。这些组织的成员或代表均是群体决策者中的一员。以群体行为做出的决策，在决策程序、决策评价标准上与单个决策者的决策有很大的差异，在决策原则、方法、许多方面都有新的内容，因而应用单个决策者的决策方法进行群体决策在许多方面都受到了限制。

群体决策(group decision)研究的是一个群体如何共同进行一项联合行动抉择。联合行动抉择一般有两种情况：一是各个决策成员参与同一行动，如公司董事会对投资项目的决策；二是各成员参与但不行动，如作为买方企业和作为卖方企业，一方是购买行动，另一方是销售行动，只有同时做出决策后，双方的行动才能付诸实施。群体决策研究的目的和单个决策者的情况一样，是描述群体决策行为的机理以及分析群体应如何进行有效的决策，

即相应分为描述性研究和规范性研究。

群体决策理论研究的问题一般具有三个前提：

（1）自主性。决策者有独立的选择机会，其行动不受较高层权力支配，但不排除群体成员的相互影响。

（2）共存性。决策成员都在已知的共同条件下进行选择。一部分成员未作出选择的情况下，其他成员的决策行动不能说最后完成。群体决策不能在撇开一部分成员的条件下去完成。

（3）共意性。群体作出的必然是所有参与者一致能接受的方案。然而，这并不意味着所有参与者都认定此方案最优。有的成员也可能持反对态度，但面临集体的最后决策而不得不作出妥协和认可。

群体中的决策问题并不都具有群体决策上述特点。企业一般属于序组织结构，下属若干车间主任，车间主任领导若干工段长等。下属的目标从上级目标派生出来并受上级的监督控制，下级服从上级，常无自主性。层序组织的领导决策实际上是个人决策。当然，各级领导在决策之前，各层次、甚至夹层次的成员也参与此决策过程，不过只是参与，最后判断和取舍则完全是领导的个人行为。自主、共存、共意并非群体决策过程的必要条件。要求所研究的群体决策问题具备上述特点，只不过是说明目前群体决策理论的局限性。



群体决策研究与个人决策研究相比，问题要复杂得多。这主要由于以下几个因素引起：

（1）偏好程度。群体的每个成员都有各自的目标和优先观念以及不同的效用函数。某些情况下成员偏好程度完全一致；而另一些情况下成员则有相互对立的偏好程度，对方的收益成为自己的受损。这是两种极端的情况。大量的情况是在群体中既有一致又有矛盾的优先观念，群体中各成员间偏好程度的矛盾强度影响着决策方式。

（2）主观概率判断。群体中各成员由于信息的感受和处理方式不一样，对未来状态出现概率的估计也不同。这直接影响着方案的选择。

（3）沟通。群体决策可以在事先完全没有沟通信息的情况下进行，在沟通过程中，相互交流各自的目标、偏好程度及对未来事件的判断，以影响对方的认识和弥补自己掌握信息的不足。

（4）人数。指群体中参与决策的人数。是两人、三人还是更多成员参与决策？这都直接影响群体决策过程的机理。一个部门、一个组织总是通过代表和其他部门和组织共同进行某项决策。因而，群体决策也研究多组织间进行的联合决策。

3.1.1 群体决策的基本概念与方法

1. 群体决策与个体决策

群体决策的理论建立在个体决策理论的基础之上，因此，个体决策理论假设也是群体决策假设，如对决策者理性的假设、偏好的传递性要求等。除此之外，群体决策由于是多个决策者共同对问题作出决策，它又有自己的特点。与个体决策比较，群体决策对问题的认知和处理等方面存在着以下的不同点。



（1）任何个体决策者都难以作出完美的决策，都可能会犯错误。这说明决策充满着风险和不确定性。

（2）至少有两名决策者需要共同负责决策。

（3）群体决策一般来说是非结构化的复杂决策问题。这说明群体决策需要解决的问题往往庞大而又复杂，单个决策者的知识和精力都有限，难以作出令人满意的决策，需要集中群体决策者集体的智慧才能创造性地解决问题。

（4）群体决策的结果应该是个体决策者的偏好形成一致或妥协之后得出的，即 Pareto 原则。这说明的是尽管决策是有风险的，但通过个体偏好的一致集结，汇集各方面的信息，又可以减少决策带来的风险和不确定性。

（5）群体决策质量受所采用的决策规则影响。

(6）群体决策质量受个体和群体的关系影响。

2. 关于群体决策的几种定义

群体决策已经成为数学、政治学、经济学、社会心理学、行为科学、管理学和决策科学等多门学科研究的共同交叉点。不同学科对群体决策研究的侧重点不同，导致形成了群体决策复杂多变的名词术语。由于群体决策问题具有内在复杂性及众多学科交叉的特性，而且研究者进行研究的角度不同，从而形成了群体决策各种各样的研究模型，也正因如此，至今群体决策也没有一种被广泛接受的统一定义。

Hwang 在 1978 年对群体决策进行分析和总结后，给出了一个群体决策的定义，即群体决策是把不同成员关于方案集合中方案的偏好按某种规则集结为决策群体的一致或妥协的群体偏好序。Hwang 的定义实际上更多地刻画出规范性群体决策的一些特征，即需要寻找一种对决策群体公平的规则对个体决策者的偏好进行集结。这个定义强调了群体决策过程是寻找每个决策个体都能够认可的群体效用函数。这个过程看起来是一个静态过程，而实际上，个体决策者在形成最终的一致或妥协的群体决策过程是一个非常复杂的过程，有可能这个决策个体意见的一致或妥协过程不得不反复进行直至决策者群体的一致偏好最终得以形成。

陈珽是这样定义群体决策的：群是由群众选出的代表组成的各种各样的委员会，群体决策是集中群中各成员的意见以形成群的意见。这个定义与 Hwang 定义比较相近。

Luce 和 Raiffa 认为群体决策问题是定义一个“公平”的方法集结个体偏好类型以至于产生由这些个体组成的社会唯一的偏好类型。能够产生这样唯一的偏好方法有很多，但并不是都是“公平”的。群体决策研究者的目的是找出这种“公平”的集结方法。由此看出，这个定义的重点是集结方法的“公平”性。

邱莞华认为：群体决策是研究多人如何作出统一的有效抉择。多个个



体组成群体，个体间可能合作，也可能竞争，还可以是复杂联合的以及合作基础上的有限竞争等，但必须作出统一的决策行为。

不同的研究者出于不同的研究视角，给出了不同的群体决策定义。

3. 群体决策的基本假设

群体决策的理论建立在个体决策理论的基础上，因此个体决策理论假设也是群体决策假设，如对决策者理性的假设、偏好的传递性要求等。群体决策由多个决策者共同对问题作出决策，它又有自己的一些特点。不同的研究者由于研究的目的不同，对群体决策研究的假设也不同。

群体决策一般存在以下基本假设。

假设 1：任何个体决策者难以作出完美决策，都可能会犯错误。假设 1 说明个体决策者在作出决策时，存在着犯错误的可能性，因此决策充满着风险和不确定性。

假设 2：至少有两名决策者需要共同负责决策。假设 2 是群体决策区别于个体决策的根本所在，由于决策者需要共同负责进行决策，决策者的个数和决策者之间的本质关系直接影响到群体决策的决策过程、决策机理以及决策结果的质量。委员会决策、组织决策以及团队决策都是由于决策者之间的关系不同而导出的群体决策形式。

假设 3：群体决策一般来说是非结构化的复杂决策问题。假设 3 指出群体决策需要解决的问题往往庞大而且复杂，单个决策者的知识和精力都极为有限，难以作出令人满意的决策，需要集中群体决策者集体的智慧才能创造性地解决问题。

假设 4：群体决策的结果应该是个体决策者的偏好形成一致或妥协之后得出的，即 Pareto 原则。由假设 1 可知，决策是有风险和不确定性的。正是通过对个体偏好的一致集结，得到来自不同来源的信息，才大大减少了决策带来的风险和不确定性。

假设 5：群体决策质量受到所采用的决策规则的影响。给定群体决策

其他因素不变，所采用的决策规则不同会得出不同的决策结果。当采用不同的决策规则时，每个备择方案都有机会成为最终的方案，深刻地说明了决策规则对群体决策质量的影响。

假设 6：群体决策质量受个体和群体关系的影响。假设 6 说明决策个体对群体的忠诚度对群体决策具有影响。

4. 群体决策的主要方法

1）机器学习法

对大量的历史数据和决策过程中积累的经验进行分析和处理以获得对决策有用的知识，主要包括：CART 学习算法、神经网路、遗传算法、粗糙集理论、基于范例的推理等。

2）软计算法

软计算法其目的在于适应现实世界普遍存在的不确定性，它是一个方法的集合。其指导原则是开拓对不精确、不确定性和部分真实的确认和表示，以达到可处理性、鲁棒性、低成本求解以及与现实更好紧密联系的目的。

3）数据仓库和联机分析处理（OLAP）

数据仓库通过多数据源信息的提取、转化、净化、汇总，建立面向主题、集成、时变、持久的数据集合，从而为决策提供依据。OLAP 是与数据仓库相关联的数据分析技术，它通过对数据仓库的即席、多维、复杂查询和综合分析，得出隐藏在数据中的事物的特征与发展规律，为决策提供支持。

4）定性推理法

定性推理理论由于其处理不完全、不确定知识和模糊数据的突出能力，在管理科学等领域受到了关注。定性推理的理论和方法被应用于预测、分析、控制和辅助决策。

这些理论和方法的运用在很大程度上突破了传统方法的局限性，提高了决策问题求解的效能和决策的智能化水平，为群体决策支持系统（Group Decision Support System，GDSS）的实现奠定了良好的方法和理论基础。



3.1.2 群体决策的协同方法

在群体决策过程中，一般先由各决策者分别做出自己的判断即评价，然后再将这些判断信息按照某种方法协同（集结）成为群体决策结果，即最终的决策。因此，群体决策过程涉及个体评价和群体决策两个阶段。关于群体决策问题的个体评价方法请读者自行查阅相关专业书籍中的多属性决策方法，下面主要总结群体决策问题的协同（集结）方法。

按照某种算法对单个评价进行集结，得到一个总体评价，称为群评价问题。群评价的集结方法也因具体问题而不同，总体上可以分为两类，即评价值的集结和评价序的集结，见表 3.1。下面分别对每种方法进行简要介绍。

表 3.1　群体决策问题的协同方法

协同类型	协同方法
基于评价值的协同	加权平均法
	算术平均法
	中间值法
基于评价序的协同	线性分配法
	平均值法
	Borda 数

1. 基于评价值的协同

设 n 个决策者分别给出对方案 i 的评价值 v_i，求群评价值 $\bar{v}$ 的算法，即为评价值的协同方法。

1）评价值协同的最优法则

对于被评价对象来讲，其本身有一个真实值或客观的合理值，评价值的最优法则应是评价值与真实值尽可能接近。

假定决策者给出的评价值 v_i，与真实值 v_0 的差是随机变量，其均值为 0，方差为 σ^2 时，且相互独立，取群评价值 $\bar{v}$ 为 v_i 的某种加权平均，即

$$\bar{v}=\sum_{j=1}^{m} w_i v_i$$

其中 $0 \leqslant w_i \leqslant 1$, $\sum_{j=1}^{m} w_i = 1$。

则$\bar{v}$也是随机变量,其均值为 v_0,方差为

$$\bar{\sigma}^2=\sum_{j=1}^{m} {w_i}^2 {\sigma_i}^2$$

要使$\bar{v}$尽可能接近 v_0,就要使$\bar{\sigma}^2$ 尽可能小,即取

$$\min \bar{\sigma}^2=\sum_{j=1}^{m} {w_i}^2 {\sigma_i}^2$$

利用该方法解此最小化问题得 $w_i=\frac{M}{{\sigma_i}^2}$,$\bar{\sigma}^2=M$,其中$M$定义为$\frac{1}{M}=\sum_{i=1}^{n}\frac{1}{\sigma^2}$。由此,评价值集结的最优法则可描述为:如果决策者 $i(i=1,2,\cdots,n)$ 的评价值为 v_i,v_i 与真实值 v_0 的差是均值为 0、方差为 σ^2 的独立随机变量,则最优的群评价值为$\bar{v}=\sum_{j=1}^{m} w_i v_i$,其中 $w_i=\frac{M}{{\sigma_i}^2}$,$\frac{1}{M}=\sum_{i=1}^{n}\frac{1}{\sigma^2}$,此时群评价值最接近真实值,误差方差最小,为$M$。

按照评价值的最优法则,评价值的协同方法主要有加权平均法、算术平均法和中间值法。

2) 加权平均法

加权平均法即与最优法相一致的方法,即

$$\bar{v}=\sum_{j=1}^{m} w_i v_i$$

其中,$w_i=\frac{M}{{\sigma_i}^2}$,$\frac{1}{M}=\sum_{i=1}^{n}\frac{1}{\sigma^2}$。

按照此方法协同的群评价值为最优值,但是应用该方法的前提是 ${\sigma_i}^2$ 已知。${\sigma_i}^2$ 反映了决策者 i 的评价水平,它是由决策者参与以往评价的历史纪录得来的。事实上,可以掌握每个决策者的评价水平的情况是极为少见的。因此实际应用中,权值的选择也不太可能按照上述公式进行。但是可以依



据此思想，为评价水平较高的决策者赋予较大的权值，为评价水平较低的决策者赋予较小的权重。

3）算术平均法

算术平均法即令

$$w_i = \frac{1}{n} \quad (i = 1, 2, \cdots, n)$$

$$\bar{v} = \frac{1}{n}\sum_{i=1}^{n} v_i$$

其评价误差方差为 $\bar{\sigma}^2 = \frac{1}{n^2}\sum_{j=1}^{n}{\sigma_i}^2$。当所有的 ${\sigma_i}^2$ 均相等时，其误差方差为$\frac{{\sigma_i}^2}{n}$达到最小。n 越大，误差方差越小。这说明当各决策者的评价水平相同或相近时，多人评价的协同比单个评价更准确，因此算术评价法可以应用于各决策者的水平相近，或者缺乏各决策者的历史评价纪录，无法区别各决策者评价水平的情况。

4）中间值法

将各决策者的评价值从大到小排序，取中间的几个值作算术平均，得到群评价值，这种协同方法称为中间值法。在竞赛评分中常用的“去掉一个（或几个）最高分，去掉一个（或几个）最低分，得平均分”的方法即为中间值法。由于最大、最小的评价值很可能是由误差方差大的决策者产生，中间值法即将他们的权重降低，而加大其余决策者的权重。因此，在各决策者评价水平不一，又无法事先知道各决策者的评价水平时，采用中间值法有利于减小总评价值的误差，特别是对于消除个别评价者有意高估或低估的影响很有效。

2. 基于评价序的协同

在某些综合评价方法特别是主观评价中，不给出各方案的评价值而直接给出各方案的优劣顺序。这种情况下多位决策者评价结果的协同就要用基于评价序的协同方法。

1）线性分配法

线性分配法实质上就是一种评价序的协同方法，不同的是它用于集结各属性的评价序以得到综合评价的排序。同样地，这种方法可用于集结各决策者的评价序以得到群评价序，这里不再详述。

2）平均值法

平均值法在各决策者给出的评价序中，对方案的位次数作简单平均，再按照此平均值排出群评价的顺序。当两方案平均值相同时可令方差较小的方案排在前面。平均值法操作简单，但它是一种较粗略的方法。

3）Borda 数

设方案集 $A=\{a_1,a_2,\cdots,a_m\}$，决策者 e_i 给出的评价序为 $u_i(i=1,2,\cdots,n)$。令 $B_i(a_j)$ 表示评价序 u_i 中后于方案 a_i 的方案个数，又令 $B(a_j)=\sum_{i=1}^{n}w_iB_i(a_j)$，其中 w_i 为决策者 e_i 的权重系数，$0\leqslant w_i\leqslant 1$，$\sum_{i=1}^{m}w_i=1$。$B(a_j)$ 称为 a_i 方案的 Borda 数。

3.1.3 德尔菲法

德尔菲法（Delphi Method）又称“专家意见法”，是为了克服专家会议法的缺点而产生的一种专家预测方法，是一种具有广泛的代表性、较为可靠且简单易行的群体决策方法。

德尔菲法依据系统的程序，采用匿名发表意见的方式，即专家之间不得互相讨论，不发生横向联系，只能与调查人员发生关系，通过多轮次调查专家对问卷所提问题的看法，经过反复征询、归纳、修改，最后汇总成专家基本一致的看法，作为预测的结果。在预测过程中专家彼此互不相识、互不往来，这就克服了在专家会议法中经常发生的专家不能充分发表意见，权威人物的意见左右其他人的意见等弊病，各位专家能真正充分地发表自己的预测意见。



1. 德尔菲法的实施步骤

德尔菲法的具体实施步骤如下：

（1）组成专家小组。按照课题所需要的知识范围，确定专家。专家人数的多少，可根据预测课题的大小和涉及面的宽窄而定，一般不超过 20 人。

（2）向所有专家提出所要预测的问题及有关要求，并附上有关这个问题的所有背景材料，同时请专家提出还需要什么材料，然后由专家做书面答复。

（3）各个专家根据他们所收到的材料，提出自己的预测意见，并说明自己是怎样利用这些材料并提出预测值的。

（4）将各位专家的第一次判断意见汇总，列成图表，进行对比，再分发给各位专家，让专家比较自己同他人的不同意见，修改自己的意见和判断。也可以把各位专家的意见加以整理，或请身份更高的其他专家加以评论，然后把这些意见再分送给各位专家，以便他们参考后修改自己的意见。

（5）将所有专家的修改意见收集起来，汇总，再次分发给各位专家，以便做第二次修改。逐轮收集意见并为专家反馈信息是德尔菲法的主要环节。收集意见和信息反馈一般要经过三、四轮。在向专家进行反馈的时候，只给出各种意见，但并不说明发表各种意见的专家的具体姓名。这一过程重复进行，直到每一个专家不再改变自己的意见为止。

（6）对专家的意见进行综合处理。

2. 德尔菲法与专家会议法的比较

德尔菲法同常见的召集专家开会、集体讨论、得出一致预测意见的专家会议法既有联系又有区别。德尔菲法能发挥专家会议法的优点：

（1）能充分发挥各位专家的作用，集思广益，准确性高。

（2）能把各位专家意见的分歧点表达出来，取各家之长，避各家之短。

同时，德尔菲法又能避免专家会议法的缺点：

（1）权威人士的意见影响他人的意见。

(2) 有些专家碍于情面,不愿意发表与其他人不同的意见。

(3) 出于自尊心而不愿意修改自己原来不全面的意见。德尔菲法的主要缺点是过程比较复杂,花费时间较长。

在这里,需要注意两个问题:

(1) 并不是所有被预测的事件都要经过步骤(1)～(4)。可能有的事件在步骤(2)就达到统一,而不必在步骤(3)中出现。

(2) 在步骤(4)结束后,专家对各事件的预测也不一定都达到统一。不统一也可以用中位数和上下四分点来作结论。事实上,总会有许多事件的预测结果都是不统一的。

德尔菲法作为一种主观、定性的方法,不仅可以用于预测领域,而且可以广泛应用于各种评价指标体系的建立和具体指标的确定过程。

3.1.4 投票表决

在群体决策的各种方法里,投票表决是在现实生活中应用最广、使用最方便、效果最明显的方法。

在实际过程中,投票表决一般由两步组成:投票和计票。投票过程应简单易行,计票过程应准确有效。根据表决过程是否进行排序,可以分为非排序式投票表决(non-ranked voting systems)和排序式投票表决(ranked voting systems)两类。下面重点对非排序式投票表决的主要情况进行归纳。

1. 只有一人当选的情况

只有一人当选时,常用的投票表决方式有计点式、简单多数制、半数代表制、二次投票法、反复投票表决法等。

1) 当候选人只有两个时

主要采用计点式(spotvote):投票采用每人一票的形式,计票采用简单多数票(simpleplurality)法则(即相对多数)。计点式是最简单的投票表决方式。





2）当候选人多于两个时

既可以采用简单多数票(相对多数)法则,也可以采用过半数(majority)法则(即绝对多数)。若采用过半数法则,当第一次投票无人获得过半数选票时,一般有两种处理方式:

(1)二次投票:对前两名进行再次投票,同候选人只有两个的情形。该投票表决方式在法国总统选举、俄罗斯总统选举中均有应用。

(2)反复投票:先淘汰部分候选人,然后重复投票过程。淘汰候选人的方式一般有两种:一是候选人自动退出,如美国两党派的总统候选人提名竞选;二是得票最少的候选人被强制淘汰,如奥运会申办城市的确定。

需要特别说明的是,无论简单多数票法则、过半数规则,还是二次投票,都有不尽合理之处。

M De Condorcet 早在 18 世纪即指出,当存在 2 个以上的候选人时,只有一种办法能严格而真实地反映群中多数成员的意愿,这就是对候选人进行成对比较。若存在某个候选人,他能按过半数决策规则击败其他所有候选人,则他被称为 Condorcet 候选人,应由此人当选。这一原则称为 Condorcet 原则。

2 两人或多人当选的情况

1）一次性非转移式投票表决(single nontransferable voting)

投票人每人一票,得票多的候选人当选。日本议员选举(选区制,每选区当选人数超过 2 个)自 1890 年起一直采用此方式。

2）累加式投票(cumulate voting)

每个投票人可投票数等于拟选出人数,选票由选举人自由支配,可投同一候选人若干票。该方式的好处在于可切实保证少数派的利益,大多用于学校董事会的选举(注意:公司董事会的选举与此不同),在英国历史上(1870—1902 年)也有应用。

3）名单制（listsystem）

由各党派团体开列候选人名单，投票人每人一票，投给党派团体，而不是直接投给候选人个人。最后根据各党派团体的名单的得票数来分配席位，并按各名单应得席位与名单上候选人的次序确定具体人选。此方式于1899年始用于比利时，以后被荷兰、丹麦、挪威和瑞典等国采用。

常用的分配席位的方法（即计票方式）有两种：最大均值法和最大余额法。可以证明，最大均值法对大党有利，最大余额法对小党有利。

3. 其他投票表决（选举）方法

下面再简单列举几种应用相对少一些的方法，因为比较容易理解，只通过简单的例子进行说明：

1）资格认定

（1）候选人数 M＝当选人数 K，即等额选举，用于不存在竞争或不允许竞争的场合。

（2）不限定入选人数，如学位点评审、职称评定、评奖等，目的不是排序，而是按某种标准来衡量被选对象。

2）非过半数规则

非过半数规则如表3.2所示。

表 3.2　非过半数规则

投票表决（选举）方法	应用案例
2/3多数	美国议会推翻总统否决需要2/3多数
2/3多数→60%多数	希腊议会总统选举，第一次需要2/3多数，第二次需要60%多数
3/4多数	美国宪法修正案需要3/4州议会的批准
过半数支持且反对票少于1/3	1993年前我国博士生导师的资格认定
一票否决	联合国安理会常任理事国的否决权

除了以上介绍的几种方法外，两人或多人当选时还有复式投票



(multiple voting)、受限的投票(limited voting)、简单可转移式选举(single transferable voting)、认可选举(approval voting)等方法。其中,复式投票是指每个投票人可投票数等于拟选出人数,且对每个候选人只能投一票;受限的投票是指每个投票人可投票数小于拟选出人数,且对每个候选人只能投一票。在实际应用中此二者均存在明显的弊端,即在激烈的党派竞争中,实力稍强的党派将拥有全部席位,因此该方法只能用于存在共同利益的团体和组织内部。

以上介绍的均为非排序式投票表决的方法。排序式投票表决的方法较非排序式复杂,其中涉及到一些非常著名、也是基础性的理论和方法,如Borda法(1770年提出)、Condorcet原则(1785年提出)、投票悖论(群的排序不具传递性,出现多数票的循环)等。除此之外,还有一些策略性投票方法,如谎报偏好、选票交易、小集团操纵、次序效应等。对于这些理论和方法,限于篇幅不在本书介绍,感兴趣的读者可自行查阅决策支持和决策分析领域的相关专业书籍。

衡量一个好的选举方法的标准应当至少具备以下三个方面的特点:

(1)能否充分利用各成员的偏好信息。

(2)若存在Condorcet候选人,应能使其当选。

(3)能防止策略性投票。

这里需要特别说明的是,目前尚没有任何一种投票表决方法对策略性投票具有防御能力。

应用案例1　人才招聘的群体决策信息协同

人才招聘决策是一个典型的不确定多属性的决策问题,需要对群体的决策进行协同。下面,以播音主持人招聘为例(5位专家、10位播音主持人),构建人才招聘的群体决策综合评价(多源决策信息协同)模型。

1. 评价指标体系建立

采用专家咨询法，筛选出能够全面描述播音员和主持人的评价指标。

构建指标体系：＜语言表达，副语言表达，职业精神，知识技能，现场表现，思想觉悟，社会调查＞。

（1）语言表达 c_1：＜语言规范度，嗓音条件，语言表现＞。

（2）副语言表达 c_2：＜化妆服饰发型，眼神表情体态＞。

（3）职业精神 c_3：＜自律能力，社会责任，专业精神，团队意识，奉献精神＞。

（4）知识技能 c_4：＜文化知识，附加技能＞。

（5）现场表现 c_5：＜参与节目的积极性，与现场人员的配合，现场效果，个人风格与节目贴合度，传播准确到位＞。

（6）思想觉悟 c_6：＜政治水平，法律意识，道德观念＞。

（7）社会调查 c_7：＜收视率，受众满意度，主持人知名度，节目美誉度＞。

2. 专家打分

（1）10 个播音主持人员构成的论域为 $U=\{x_1,x_2,x_3,\cdots,x_{10}\}$。7 个一级指标构成了条件属性集 $C=\{c_1,c_2,c_3,c_4,c_5,c_6,c_7\}$。评价等级{优，中，差}分别对应{2，1，0}。

（2）5 个专家根据 7 个指标和 3 个评价等级，分别对 10 个播音主持人员的表现进行打分。设定 5 个专家的意见是同等重要的，因此取 5 个专家的平均分并取整，系统计算出 10×7 关系矩阵，如表 3.3 所示。

表 3.3　关系矩阵

	c_1	c_2	c_3	c_4	c_5	c_6	c_7
x_1	2	1	2	1	1	1	1
x_2	2	2	1	1	2	1	2
x_3	1	1	2	1	1	2	1
x_4	1	2	1	2	2	1	2
x_5	0	0	1	1	0	1	0
x_6	2	2	0	1	2	0	2

续表

	c_1	c_2	c_3	c_4	c_5	c_6	c_7
x_7	0	1	0	1	1	0	1
x_8	1	0	1	1	0	1	0
x_9	1	2	0	1	2	0	2
x_{10}	1	1	1	1	2	1	2

3. 指标体系约简

系统依据 Skowron 提出的信息系统区分矩阵的定义和表 3.3 构造出区分矩阵，如表 3.4 所示。

表 3.4　区分矩阵

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
x_1	0	0	0	0	0	0	0	0	0
x_2	c_2,c_3,c_5,c_7	0	0	0	0	0	0	0	0
x_3	c_1,c_6	c_1,c_2,c_3,c_5,c_6,c_7	0	0	0	0	0	0	0
x_4	c_1,c_2,c_3,c_4,c_5,c_7	c_1,c_4	c_2,c_3,c_4,c_5,c_6,c_7	0	0	0	0	0	0
x_5	c_1,c_2,c_3,c_5,c_7	c_1,c_2,c_5,c_7	c_1,c_2,c_3,c_5,c_6,c_7	c_1,c_2,c_4,c_5,c_7	0	0	0	0	0
x_6	c_2,c_3,c_5,c_6,c_7	c_3,c_6	c_1,c_2,c_3,c_5,c_6,c_7	c_1,c_3,c_4,c_6	c_1,c_2,c_3,c_5,c_6,c_7	0	0	0	0
x_7	c_1,c_3,c_6	c_1,c_2,c_3,c_5,c_6,c_7	c_1,c_3,c_6	$c_1,c_2,c_3,c_4,c_5,c_6,c_7$	c_2,c_3,c_5,c_6,c_7	c_1,c_2,c_5,c_7	0	0	0
x_8	c_1,c_2,c_3,c_5,c_7	c_1,c_2,c_5,c_7	c_2,c_3,c_5,c_6,c_7	c_2,c_4,c_5,c_7	c_1	c_1,c_2,c_3,c_5,c_6,c_7	c_1,c_2,c_3,c_5,c_6,c_7	0	0
x_9	c_1,c_2,c_3,c_5,c_6,c_7	c_1,c_3,c_6	c_2,c_3,c_5,c_6,c_7	c_3,c_4,c_6	c_1,c_2,c_3,c_5,c_6,c_7	c_1	c_1,c_2,c_5,c_7	c_2,c_3,c_5,c_6,c_7	0
x_{10}	c_1,c_3,c_5,c_7	c_1,c_2	c_3,c_5,c_6,c_7	c_2,c_4	c_1,c_2,c_5,c_7	c_1,c_2,c_3,c_6	c_1,c_3,c_5,c_6,c_7	c_2,c_5,c_7	c_2,c_3,c_6

根据区分矩阵构造出区分函数并用吸收率进行化简得：

$$f=(c_1 \wedge c_2 \wedge c_3) \vee (c_1 \wedge c_3 \wedge c_4 \wedge c_5) \vee (c_1 \wedge c_3 \wedge c_4 \wedge c_7)$$
$$\vee (c_1 \wedge c_2 \wedge c_6) \vee (c_1 \wedge c_4 \wedge c_5 \wedge c_6) \vee (c_1 \wedge c_4 \wedge c_6 \wedge c_7)$$

从区别矩阵中可以看出指标集的核是 c_1，即语言表达，说明基本功对主持人的重要性。这些约简是并列关系，可以选取任意一个作为约简结果，一般选择指标个数最少的约简集。在此，专家根据系统计算出的约简集，并结合对评价对象的要求进行选择，这里选取 $\{c_1, c_2, c_6\}$（语言表达，副语言表达，思想觉悟）。用这种方法有效地缩减了评价指标的规模。

最终确定的指标体系为：

(1)语言表达 c_1：＜语言规范度，嗓音条件，语言表现＞。

(2)副语言表达 c_2：＜化妆服饰发型，眼神表情体态＞。

(3)思想觉悟 c_6：＜政治水平，法律意识，道德观念＞。

4. 指标权重确定

根据对播音主持人评价构成的信息系统表可求得约简后的指标权重：

C/D 的等价类：$\{\{x_1\},\{x_2\},\{x_3\},\{x_4\},\{x_5\},\{x_6\},\{x_7\},\{x_8\},\{x_9\},\{x_{10}\}\}$。

$C\text{-}\{c_1\}$的等价类：$\{\{x_1,x_{10}\},\{x_2,x_4\},\{x_3\},\{x_5,x_8\},\{x_6,x_9\},\{x_7\}\}$。

$C\text{-}\{c_2\}$的等价类：$\{\{x_1,x_2\},\{x_3\},\{x_4,x_8,x_{10}\},\{x_5\},\{x_6\},\{x_7,x_9\}\}$。

$C\text{-}\{c_6\}$的等价类：$\{\{x_1\},\{x_2,x_6\},\{x_3,x_7,x_{10}\},\{x_4,x_9\},\{x_5\},\{x_8\}\}$。

系统根据公式计算指标的权重如下：

$$I(C)\sum_{i=1}^{10}\frac{|x_i|}{|U|}\left(1-\frac{|x_i|}{|U|}\right)=\frac{1}{10}\times\left(1-\frac{1}{10}\right)\times 10=\frac{9}{10}$$

$$I(C-\{c_1\})=\frac{2}{10}\times\left(1-\frac{2}{10}\right)\times 4+\frac{1}{10}\left(1-\frac{1}{10}\right)\times 2=\frac{82}{100}$$

同理 $I(C-\{c_2\})=\dfrac{80}{100}$，$I(C-\{c_6\})=\dfrac{80}{100}$。

$$w_{c_1}=\frac{\mathrm{sig}_{A-\{c_1\}}}{\sum\limits_{j=1}^{n}\mathrm{sig}_{A-\{c_j\}}(c_j)}=\frac{I(C)-I(C-\{c_1\})}{nI(C)-\sum\limits_{j=1}^{n}I(C-\{c_j\})}$$



$$=\frac{\frac{9}{10}-\frac{82}{100}}{3\times\frac{9}{10}-\left(\frac{82}{100}+\frac{80}{100}+\frac{80}{100}\right)}=\frac{2}{7}$$

同理 $w_{c_2}=\frac{5}{14}$，$w_{c_6}=\frac{5}{14}$。

5. 综合评价协同

根据指标权重和约简后的指标值，采用线性加权法对各个指标进行加权计算，求得 10 个播音主持人的综合评价得分：

$$(0.2857,0.3571,0.3571)\times\begin{bmatrix}2&2&1&1&0&2&0&1&1&1\\1&2&1&2&0&2&1&0&2&1\\1&1&2&1&1&0&0&1&0&1\end{bmatrix}$$

$$=(1.29\quad 1.64\quad 1.36\quad 1.36\quad 0.36\quad 1.29\quad 0.36\quad 0.64\quad 1.00\quad 1.00)$$

10 名播音主持人的排序结果为：

$$x_2>x_3=x_4>x_1=x_6>x_9=x_{10}>x_8>x_5=x_7$$

采用 TOPSIS 方法计算的步骤在此不详述，10 名播音主持人的排序结果为：

$$x_3>x_2>x_4>x_1>x_{10}>x_6>x_9>x_8>x_5>x_7$$

可以看出，线性加权法和 TOPSIS 法得到的排序结果基本一致，但在个别主持人的排序上因选择方法的不同产生了微调。

3.2 面向城市管理的多源信息协同

3.2.1 信息协同的内涵与前沿趋势

1. 信息协同的内涵与基本概念

德国斯图加特大学教授、著名物理学家哈肯（Hermann Haken）于 1971 年提出了“协同”的概念，并于 1976 年系统地论述了协同理论，主要研究远离

平衡态的开放系统在与外界有物质或能量交换的情况下，如何通过自己内部协同作用，自发地出现时间、空间和功能上的有序结构。

信息协同是指以信息为对象，多个信息源在规定的时间和空间内，按照统一的规则实现信息的有序流转。与信息协同相关的基本概念主要有信息提供方、信息需求方、信息协同节点、信息协同流程等。

（1）信息提供方：提供信息的组织（或个人），信息协同流程的发起方。

（2）信息需求方：获取信息的组织（或个人），信息协同流程的接收方。

（3）信息协同节点：位于信息组织（或个人）内部，实现与其他信息组织（或个人）之间的信息协同流程。在管理定位上相当于“传达室”的功能，在技术层面一般表现为前置机的形式。

（4）信息协同流程：每一类信息在信息提供方和信息需求方之间的共享对应一个信息协同流程，在业务和技术层面分别对应信息协同业务流程和信息协同技术流程。

与传统城市运行中的信息获取方式不同，智慧城市运行中的信息包含大量的物联信息，来自于对城市中不同实体状态和行为的实时感知，呈现出多源、异构、海量、实时、不间断的显著特征。物联信息正逐步取代传统的业务信息，成为城市运行管理中信息流转的主体。



物联信息主要包括实时感知信息和物联综合信息两类。

（1）实时感知信息：指通过感知设备实时采集或读取的信息，如自动气象站温度实时信息、消防车辆 GPS 实时信息等。

（2）物联综合信息：指在智慧城市领域应用中通过汇总、分析等方式形成的综合性信息，如某地区的天气预报信息、某区域的危险化学品流向信息等。

物联信息通过物联信息元数据来描述具体的内容、质量、表示方式、管理方式以及数据集的其他特征。物联信息元数据由元数据实体和元数据元素组成，是实现物联信息共享的核心部分。

物联信息所涉及的实体统称为物联实体，主要包括感知设备、管理对象

和传感网网关三类。

(1) 感知设备(sensing device):指能够实时监测、感受、识别外界信息,并将获取的信息传递给其他装置的物理设备,如温度传感器、数字摄像头、卫星定位设备、射频识别读写器等。

(2) 管理对象(managed object):指需要通过感知设备来进行辅助监管和监控的业务实体对象,主要包括地理实体、物品和证照三类。地理实体包括自然地理实体(如山川、河流、湖泊、土地等)和人工建造的地理实体(如建筑、街道、公路、桥梁、雕塑、渠坝、场站等);物品包括除了地理实体之外可以移动的实体,如车辆、井盖、感知设备、计算机等;证照包括由政府部分发放的具有法定效力的证件,如身份证、营业执照等。

(3) 传感网网关(wireless sensor networks gateway):指能够将物联感知设备采集的实时感知信息进行接入和处理,并按照规范规定的格式和内容打包后依托网络设施进行传输的一种设备。传感网网关通过用户前置设备(customer premise equipment,CPE)与物联数据专网进行连接。



2. 信息协同的前沿趋势

1) 国外研究热点和趋势

根据 Web of Science-SCI 和 Engineering Village(EI)的收录情况显示,截至 2014 年年底国外与信息协同(information cooperation)相关的研究成果共 1131 篇,其中 SCI 312 篇、EI 819 篇,呈现出明显的逐年递增趋势。研究人员和成果的分布区域以中国和美国较为突出,此外日本、德国、法国、加拿大、英国、意大利、韩国、澳大利亚也有较高占比。

SCI 的研究领域中自然科学约占 63%,社会科学约占 17%,自然科学与社会科学的交叉领域约占 20%,研究方向以计算机科学、信息科学、工程学为主(见表 3.5)。

EI 的研究成果所涉及学科分布相对比较平均,在计算机软件和信息处理、计算机应用、通信、信息传播、数据处理和图像处理、信息理论和信号处理、

表 3.5 信息协同的 SCI 研究方向

SCI 研究方向	数量	占比
Computer Science 计算机科学	96	30.77%
Information Science Library Science 信息科学、图书馆学	65	20.83%
Engineering 工程学	51	16.35%
Business Economics 商学、经济学	35	11.22%
Psychology 心理学	29	9.29%
Health Care Sciences Services 健康护理学、服务科学	28	8.97%
Telecommunications 电信学	21	6.73%
International Relations 国际关系学	21	6.73%
Communication 传播学	16	5.13%
Operations Research Management Science 运筹学、管理科学	14	4.49%

信息检索和利用、数学、人工智能等领域均有代表性文献；研究热点集中在信息系统、信息传播、多主体系统、算法、计算机模拟、信息理论、信息检索、信息管理、最优化、信息分析等领域（见表 3.6）。



表 3.6 信息协同的 EI 研究热点与学科分布

（a）EI 研究热点

EI 研究热点	数量	占比
Information Systems 信息系统	81	9.89%
Information Dissemination 信息传播	74	9.04%
Multi Agent Systems 多主体系统	56	6.84%
Algorithms 算法	52	6.35%
Computer Simulation 计算机模拟	51	6.23%
Information Theory 信息理论	49	5.98%
Information Retrieval 信息检索	48	5.86%
Information Management 信息管理	47	5.74%
Optimization 最优化	44	5.37%
Information Analysis 信息分析	43	5.25%

续表

(b) EI学科类别

EI学科类别	数量	占比
Computer Software，Data Handling and Applications 计算机软件、数据处理与应用	244	29.79%
Computer Applications 计算机应用	207	25.27%
Telecommunication；Radar，Radio and Television 通信；雷达、收音机和电视	207	25.27%
Information Dissemination 信息传播	182	22.22%
Data Processing and Image Processing 数据处理和图像处理	128	15.63%
Information Theory and Signal Processing 信息理论和信号处理	120	14.65%
Information Retrieval and Use 信息检索和利用	111	13.55%
Mathematics 数学	110	13.43%
Artificial Intelligence 人工智能	104	12.70%
Radio Systems and Equipment 无线电系统和装备	100	12.21%

2）国内研究热点和趋势

根据SCI来源期刊、EI来源期刊、核心期刊(北京大学2011版)、CSSCI期刊的收录情况显示，截至2014年年底国内关于“信息协同”的理论和应用研究共计441篇相关文献。2012年(含)之后的研究成果数量及国家自然科学基金、国家高技术研究发展计划(863计划)、国家科技支撑计划、国家社会科学基金等的支持项目数量均有明显的提升。

研究层次主要分布在工程技术(自然科学)、行业指导(社会科学)、基础研究(社会科学)、基础与应用基础研究(自然科学)、行业技术指导(自然科学)、政策研究(社会科学)等方面。研究热点主要集中在信息协同模式与模型、行业应用、协同管理、服务与创新、信息共享与融合、信息服务与信息化、信息系统与Web服务、知识管理、协同机制与模式、物联网、协同过滤、云计算等领域(见表3.7)。

表 3.7　信息协同的国内研究热点

国内研究热点	数量	占比	国内研究热点	数量	占比
信息协同模式与模型	57	18.75%	信息系统与 Web 服务	18	5.92%
行业应用	55	18.09%	知识管理	15	4.93%
协同管理、服务与创新	52	17.11%	协同机制与模式	14	4.61%
信息共享与融合	20	6.58%	物联网	11	3.62%
信息服务与信息化	18	5.92%	协同过滤	10	3.29%

3.2.2　面向城市管理的信息协同应用模式

1. 多源信息协同的业务模式

面向城市管理的多源信息协同业务模式主要包括单部门为主多部门配合、多部门流程化信息协同、多部门交叉信息协同、多源信息整合、基础信息协同等五种情况。

（1）单部门为主多部门配合模式：某项事件涉及多个部门，但以一个部门为主，其他相关部门配合开展事件处理。一个部门负责信息协同流程的发起，将信息分别发送给多个相关部门进行协同处理，结合各部门反馈的结果信息进行事件处理。

（2）多部门流程化信息协同模式：某项事件涉及多个部门，相互之间通过前后置协作实现联合处理。一个部门负责信息协同流程的发起，前置环节结果信息作为后置环节事件处理的必要信息。

（3）多部门交叉信息协同模式：两个及以上部门对同一事件同时进行处理，或侧重同一事件的不同方面，通过信息协同实现各个环节的及时、有效开展。多部门同时处理同一事件时，部门间通过内部协同机制实现信息联动。多部门处理同一事件但侧重领域或面向对象不同时，各部门分别进行事件处理，通过信息协同获取其他部门信息作为事件处理的参考和依据。

（4）多源信息整合模式：由一个部门统一接入多个部门的信息，整合融



合后提供信息服务。一个部门接入获取多个部门的信息，进行加工整合、分析汇总、多维融合，形成综合性、主题性信息，以统一出口发布。

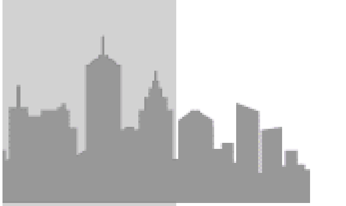
(5) 共性基础信息协同模式：根据实际需求，各部门或信息组织将有共性需求的基础信息(如人口基础信息、法人基础信息、空间地理基础信息、宏观经济基础信息、城市部件基础信息、城市基础运行状态基础信息等)统一汇聚、整合和服务。

2. 多源信息协同的技术模式

面向城市管理的多源信息协同技术模式主要包括信息交换、信息目录、接口调用等多种情况。

(1) 信息交换模式：根据跨部门、跨区域信息共享需求，部门之间协商确定信息共享内容、交换频率、提供部门、需求部门等，建立信息交换对子。需求部门按照交换规则获取提供部门的各类信息。

(2) 信息目录模式：提供部门对可协同信息进行编目，形成目录信息，在统一的基础支撑平台上进行注册；需求部门通过查询信息目录，访问可协同信息的目录信息，并依据权限访问具体信息内容。



(3) 接口调用模式：提供部门对可协同信息进行封装形成服务，在统一的基础支撑平台上进行注册，形成服务清单；需求部门查询服务清单，访问数据库、页面、Portlet、Web 服务、应用系统等服务资源的描述信息，并依据权限访问、调用服务资源接口。

除此之外，还有页面集成整合、基于业务协同的信息审核确认等其他技术模式。

3. 多源信息协同的推进模式

面向城市管理的多源信息协同推进模式主要包括：重大应用推进模式、重大热点问题协同推进模式、部门整体共享推进模式(某个部门把需要内外共享的信息整体协调推进)、决策服务推进模式、业务协同推进模式(业务领域/业务主题相关部门之间的协同)、常规结对子共享推进模式等。

4. 多源信息协同模式的其他维度分析

除了业务模式、技术模式、推进模式等维度的研究之外，面向城市管理的多源信息协同还有以下一些研究维度：

（1）从层次维度：同级组织间信息协同、跨层级信息协同。

（2）从范围维度：信息组织内部、信息组织之间。

（3）从信息分布维度：集中共享服务、分布式交换等。

（4）从信息性质维度：基础信息、领域主题信息等。

5. 多源信息协同水平的测度

目前，面向城市管理的多源信息协同水平测度主要包括五个方面：一是目录更新情况，是否按计划进行需求目录和共享目录的更新或审核确认；二是被需求目录的响应情况，是否及时响应被需求目录的情况；三是信息提供情况，是否按计划进行数据提交和更新，提交数据的内容与目录是否一致；四是数据质量情况，向其他部门提供的数据质量是否满足需求；五是应用案例，信息协同是否支撑了重大应用或有典型的应用案例。



3.2.3 多源信息融合的内涵与发展

1. 多源信息融合的内涵

1）从信息环境看多源信息融合

当前的数据环境已呈现出从海量数据环境向大数据环境转化的趋势。大数据的特点之一是数据类型繁多，基于各种数据类型的分析是大数据的典型特点之一。电子邮件、访问日志、交易记录、社交网络、即时消息、视频、照片、语音等，是大数据的常见形态，这些数据从不同视角反映人物、事件或活动的相关信息，把这些数据融合汇聚在一起进行相关分析，可以更全面地揭示事物联系，挖掘新的模式与关系，从而为科学有效的管理决策、商业模式的制定、竞争机会的选择提供有力的数据支撑与决策参考。

同一个事实或规律可以同时隐藏在不同的数据形式中，也可能是每一

种数据形式分别支持了同一个事实或规律的某一个或几个侧面，这既为数据和信息分析的结论的交叉验证提供了契机，又要求分析者在分析研究过程中有意识地融集各种类型的数据，从多种信息源中发现潜在知识与情报。因此，综合利用多来源、多形式的数据是现代科学决策的鲜明特点。“兼听则明，偏信则暗”，多维度、多数据源的分析才有说服力。

2）从业务需求看多源信息融合

通过关联主题分析可以识别出一些模式，从而实现需求驱动的跨业务的关联分析。例如，根据常住人口数据、流动人员数据、房屋数据，运用关联规则挖掘方法分析城乡结合部流动人口密集与多发事件的关系，通过对海量数据的统计分析、时间序列、趋势外推、监测分析、价值挖掘、场景预测等进行模式发现、规律验证、趋势预测、根源分析、舆情监控等，从而实现事前预防、综合应急、科学决策以及支撑快速处理等。实现多渠道信息汇聚“一口进”，决策指令才能“一口出”，减少信息错漏，防止决策失误。



以现有的数字化城市运行体征情况为基础，建立科学的预警机制，实现对城市管理问题科学、准确地事前预防与快速处理，依据城市运行以及问题发生的规律、趋势、根源的分析，依据城市运行动态数据为领导提供强大的数据支撑和多视角的决策分析。

3）多源信息融合的基本概念

把不同渠道、利用多种采集方式获取的具有不同数据结构的信息汇聚到一起，形成具有统一格式、面向多种应用的数据集合，这一过程称为多源信息融合。多源信息融合技术是研究如何加工、协同利用多源信息，并使不同形式的信息相互补充，以获得对同一事物或目标的更客观、更本质认识的信息综合处理技术。它比直接从各信息源得到的信息更简洁、更少冗余、更有用途。传统的数据融合是指对多传感器的数据在一定准则下加以自动分析、综合的信息处理过程。

4）多源信息融合的类型

多源信息融合包括同型异源信息的融合、异质异构信息的融合、多语种

信息的融合。同型异源信息融合是指同一种类型的信息可能分布在不同的物理存储、不同的应用系统，隶属于多个机构部门，这些不同来源的信息有着不同的来源渠道、采集方式、加工体系与标准，也有着不同的服务模式，使用单一数据源进行分析很难保证全面性，实际分析时往往需要将多种数据集成到一起进行融合分析。异质异构信息融合是指把不同类型不同结构的信息汇聚到一起，以更全面地反映现状，说明问题。例如在学科领域分析时需要把期刊论文、学位论文、图书、专利、项目、会议等信息融合汇总，重大事件监测与分析时需要把数值、文本、音频、视频等不同媒体格式的信息汇聚到一起进行关联综合分析。多语种信息融合是指把不同语种的信息融合汇聚到一起进行综合集成分析，以提高信息的全面性。

2 多源信息融合的发展现状

1）信息融合的基础理论与架构

信息融合理论最早应用于军事领域，定义为一个处理探测、互联、估计以及组合多源信息和数据的多层次、多方面的过程，以便获得准确的状态和身份估计、完整而及时的战场态势和威胁估计。而随着网络海量信息化的进程，信息融合逐渐得到更多领域的关注。朱子华等根据对图书复杂适应系统的研究，分析了图书信息融合系统的功能需求、体系结构层次和技术结构层次，认为信息的融合主要体现在不同角色之间的信息交互。胡蓓等提出产业集群知识融合的新观点：应用理论研究与实证研究相结合的方法，采用信息融合的 DSmT 技术，对集群内、外部知识源的知识进行融合。刘明香将基于 D-S 证据推理的信息融合技术应用于知识转化为信息的过程中，使人们更准确快捷地获取信息。

2）信息融合的层次体系

信息融合是在几个层次上完成对多源信息的处理过程，其中每一层次都表示不同级别的信息抽象；信息融合的结果包括较低层次上的状态和身份估计，以及较高层次上的整个战术态势估计。曹建君等把信息融合划分

为原始数据融合或象元级融合、目标级或特征级融合以及决策级融合3个层次。任红娟从数据库扩展的原始级信息融合、文本挖掘和文献计量方法结合、词汇引用图和词参考文献共4个层面对当前的知识结构整合方法进行了综述。宋新平等针对竞争情报系统循环的不足以及情报源的特点，构建一个基于信息融合综合集成研讨厅混合的新型竞争情报系统。

3）信息融合的技术与流程

刘平峰等针对现有Web信息融合对多维度、多粒度综合查询分析和决策支持不足的问题，设计面向主题的Web信息融合模型，该模型由Web仓库模型、Web信息融合功能模型和人机交互接口三层组成。陈金海针对目前科研成果中信息融合的特点，研究了情报信息融合处理方法的多样性，其中主要包括情报信息分类处理过程、科技信息融合处理和情报数据的融合技术处理过程等。

4）多源信息融合的关联分析



多源信息融合的理论基础是相关性原理。相关性分析包括主题相关、任务相关、情境相关、用户相关等。相关关系的核心是量化两个数据值之间的数理关系。相关关系强是指当一个数据值增加时，另一个数据值很有可能也会随之增加。例如谷歌流感趋势：在一个特定的地理位置，通过谷歌搜索流感相关的特定词条的人越多，表明该地区患流感的人越多。不同来源的数据可以从不同视角反映人物、事件或活动的相关信息，把这些数据融合汇聚在一起进行相关分析，可以更全面地揭示事物联系，挖掘新的模式与关系，从而为科学有效的管理决策、商业模式的制定、竞争机会的选择提供有力的数据支撑与决策参考。

3. 多源信息融合的业务需求

以现有的数字化城市管理情况为基础，建立科学的预警机制，实现城市管理问题科学、准确地事前预防与快速处理，实现城市运行以及问题发生的规律、趋势、根源的分析，通过城市运行动态数据为城市管理者提供强大的

数据支撑和多视角的决策分析。通过对海量数据的统计分析、趋势外推、监测分析、价值挖掘、场景预测等进行模式发现、规律验证、趋势预测、根源分析、舆情监控等，实现事前预防、综合应急、科学决策以及支撑快速处理等。从决策与管理需求角度来看，多源信息融合的业务需求主要包括综合应急、舆情监控、预警监测、问题定位、形势分析、模式分析等。

1）综合应急

应急数据综合汇聚，当有重大突发事件时，把事件所涉及的人员、房屋、城市部件、应急物资等按地理半径或影响程度进行汇聚呈现，以实现重特大突发事件应对过程中的信息快速交换、精确传播、主辅责部门之间的横向联动和协同作战，使应对和处置的效率大幅提升，快速反应、协同应对、合成应急能力显著提升。

2）舆情监控

通过热线电话、网络在线、网格员上报、突发事件等实时汇聚热点问题，把握舆情动态与走向，包括民众关心的热点问题、观点倾向、舆情演化动态等。

3）预警监测

在业务数据基础上实现趋势预测分析，实现专项数据的同比、环比得到若干年度的趋势图，并以可信的变化率（例如专业机构、主管机构公开发布的数据）为基准对未来若干年度做出预测。通过“六位一体”信息汇聚网络及网格力量反馈，构建多渠道、全覆盖的预警监测体系，实时掌握风险隐患和动态变化，破解预警传播“最后一公里”的难题。

4）问题定位

通过城市管理各类监管案件、公众及企业诉求热点等相关指标比较，通过量化分析来定位问题，以便及时查找问题、分析问题原因并制定相关措施。

5）形势分析

加强数据分析和案例剖析，形成城市管理监督指挥、网格化社会服务管



理、突发事件的“日统计、周汇总、月分析和季度分析”机制，为部署下阶段工作重点、提出具体应对之策提供翔实依据。

6）模式分析

通过关联式主题分析识别出一些模式，实现需求驱动的跨业务的关联分析。如通过对实有人口与市容环境、日常隐患与突发事件、地理空间实有人口与突发事件等几个方面的资源关联分析，实现对城市运行管理问题科学、准确、实时地事前预防与快速处理。

4. 多源信息融合的发展方向

1）业务信息与空间信息的融合

多源数据融合中主要涉及业务数据与空间数据，在建设智慧城市的进程中，空间数据的建设、分析与应用得到了前所未有的关注。应用多维度、多时空的全时空一体化技术，建设时空多维城市管理系统，将二维地图、三维模型、三维实景、视频等各类基础信息与城市部件、实有人口、法人、房屋等行业管理信息整合到一个时空体系中，在全时空体系中还原各类信息在空间、时间、权属等多个维度的关联关系，创新三维立体直观的、实时的可视化管理模式，提高社会管理、城市管理、社会服务、应急管理工作的效率。



2）多源信息的统一描述与深度揭示

不同来源不同结构的数据融合到一起，如何对数据资源进行统一描述、深度揭示是数据融合的关键问题之一。虽然实现了多种来源、多种采集方式的数据融合，但是对各类文本信息、热线电话录音、各种监控视频等缺乏统一描述，深度揭示。统一描述有助于组织、交换与融合，深度揭示有助于进行分析与挖掘，提高数据的利用价值，因此需要从内容上对多种信息来源不同数据格式的信息进行融合。

3）多源信息的动态关联与交叉印证

城管平台运行相关数据、非紧急救助热线相关数据、社区台账相关数据、网格上报相关数据、数据容器抓取相关数据、应急指挥平台相关数据、部

门提供相关数据等多源数据在实现初步融合的基础上，进一步形成系统、完整的信息链条，各口径数据均分类存储在不同的专题数据库中，而单一的基础数据库在解决综合性问题的时候能力不足，迫切需要各类数据的全面融合来支撑复杂的、涉及范围较广的问题的解决。结合城市管理的实际业务，将来在数据融合方面，需要对各种数据进行动态关联、交叉印证，确保数据的鲜活性与统一性，为实现城市管理提供客观、全面、准确的数据支撑。

3.3 智慧城市多源信息协同体系的理论框架

3.3.1 当前智慧城市建设存在的主要问题

目前，国内外的智慧城市建设已初具规模，相关技术和产业发展迅速，信息化对城市化的引领和促进效应凸显。但是，关于智慧城市的理论研究还不成熟，没有形成完整的理论体系，主要存在三个方面的问题：一是顶层设计和总体规划层面缺乏科学的方法论指导，二是建设模式与实施路径上缺乏成功的案例，三是在大部分技术环节和应用领域中尚未形成统一的标准。

国外的智慧城市建设具有明显的地区特色，单一领域的应用较为深入和成熟，但没有形成通用性的、规模性的智慧应用和服务。国内的情况则恰恰相反，主要表现出三个特点：一是大多数智慧城市的规划比较相近，在设计上“大”而“全”，缺少地区特色的结合和地域间的差异性；二是部分智慧城市的建设过于依赖 IBM 等国外企业的支撑。由于国外新一代信息技术的应用发展和智慧城市建设本身也在探索过程之中，相关理论和技术均不成熟，且国内外的国情和体制完全不同，面临的社会问题重点不同，信息化和城市化的发展思路与环境也有较大差别，因此完全照搬国外智慧城市的建设模式并不合理；三是国内的智慧城市建设目前多数仍然停留在数字城市阶段，不乏大量的“面子工程”和“政绩工程”，虽然举着“智慧城市”的大旗，实际上做的都是在“数字城市”阶段早就应该解决的问题，许多城市部门间的信息

共享渠道仍然没有打通，甚至有的地区尚未实现办公文档的电子化，也跟风出台了一系列智慧城市的发展规划，存在明显的“大跃进”现象。

从总体上看，国内外关于城市信息化的发展方向主要分为三个方面：一是从城市地理或地理信息系统(GIS)的角度，研究信息化背景下的城市空间结构变化，以及城市空间信息系统及其体系结构和技术架构；二是从城市规划管理的角度，研究在城市规划、建设、管理与服务中的信息化应用与运行模式；三是从城市社会经济发展的角度，研究信息化对城市社会经济生活各个方面的影响、带动作用及如何有效应用信息技术和信息服务促进城市的全面发展。通过分析可以发现，在城市空间信息系统的体系架构、城市规划层面的信息化应用、面向城市经济发展和社会管理的信息技术应用和信息服务中均涉及多源信息的协同和联动，信息协同问题是城市信息化发展的共性问题，也是最核心的环节。



目前，国内外的信息协同技术发展比较成熟，但在城市和区域层面的信息协同应用相对较少，尚未形成有代表性的案例和模式。主要存在以下几个方面的问题：

(1) 没有建立起统一的信息协同标准体系框架，且在关键的共性问题(如实时信息的分类、基础信息的属性和编码、多源信息的接入与传输等)上没有形成统一的标准。智慧城市区别于数字城市和传统城市的突出特点在于大量存在的传感器和移动设备及由感知端和移动端产生的海量实时信息，对基础信息和实时信息的多维度、精细化管理是智慧城市管理的重要基础。目前，国内外在信息共享和传输等技术层面存在多种国际标准、国家标准和行业标准，在市政管理、交通管理、公安、水务、安全生产等城市基础运行管理的不同领域形成了传感器类别编码标准等实操性较强的地方标准和行业标准，但是上述标准的领域性普遍较强，没有形成基于业务特点的共性需求，在城市和区域层面上尚不具有通用性。

(2) 城市系统下的信息协同模式不能适应智慧城市大数据决策和精细化管理的需求。目前城市系统下的信息流转模式比较单一(以中心主导的

管理模式为主)，感知信息、事件信息和决策信息的流向相对固定，一方面不足以应对秒级(甚至毫秒级)的海量物联信息产生和流转频率，另一方面也造成了信息资源的重复流转和多次整合，协同效率低下。同时，面向智慧城市大数据特征的信息协同总体架构和关键技术需要重新定义，以便能够灵活地支撑多种信息协同模式的自适应过程，以应对智慧城市管理对快速决策、应急指挥和精细化管理提出的新的需求。

(3) 作为具有导向作用的城市信息化测度体系方面的发展相对滞后，应用主要集中在总体层面，缺少对结构层面的分析。总体层面的测度主要关注总量指标，如信息共享目录和信息需求目录的变化情况、信息节点接入信息交换技术通道的情况、信息协同的需求响应情况、信息协同技术流程与业务流程的关联情况等，测度的重点是信息协同的及时性(需求响应度)、协同信息的质量(信息的准确性、连续性、匹配度)和多源信息之间的关联度。而结构层面的测度主要关注个体及其之间的关系，重点分析城市系统下信息协同网络结构中不同信息组织的差异性和信息组织之间的关系。总体测度与结构测度的关系可以类比为某个国家或地区的 GDP 水平与基尼系数之间的关系。

(4) 城市系统下的多源信息融合应用尚处于初级阶段。一方面，信息融合在传感器领域研究相对较多，在信息管理领域研究不足，只有需求比较单一的信息关联整合应用，缺少结合情景的、事件驱动的复杂实证研究；另一方面，信息组织内部的信息融合居多，跨部门、跨层级、跨领域的信息融合较少，且缺少对信息融合技术层面与建模方案的深入分析。

3.3.2 智慧城市多源信息协同的理论模型

面向城市和区域层面的多源信息协同体系重点探讨信息在城市系统中的协同模式及其相关的标准、机制、技术、模型和策略，从信息流的角度探讨信息化对城市运行和精细化管理的提升作用。智慧城市多源信息协同的理论模型如图 3.1 所示。



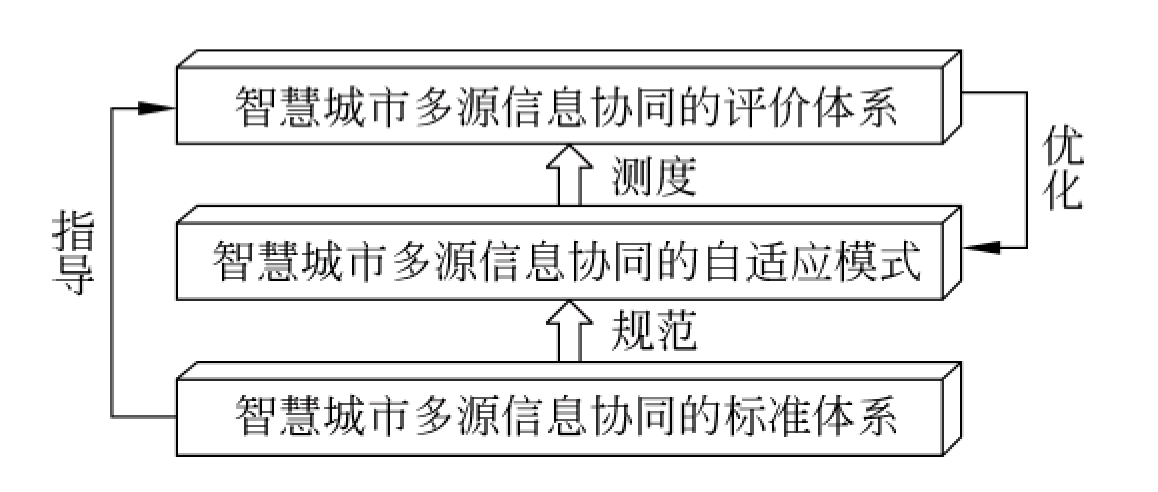


图 3.1　智慧城市多源信息协同的理论模型

由图 3.1 可知，智慧城市多源信息协同体系主要由信息协同标准、信息协同模式、信息协同评价三个层次构成。其中，信息协同标准是信息协同模式的共性规范和约束，同时指导信息协同的评价，即信息协同网络的测度过程；信息协同模式与信息协同评价之间形成城市系统下信息流转的闭环流程，根据信息协同网络的测度结果进行信息协同模式的智能优化。

智慧城市多源信息协同体系的核心环节主要包括三个方面：

（1）信息协同的标准体系框架及其关键标准：通过标准体系将智慧城市运行管理所涉及的人、地、事、物、组织用信息关联起来，实现信息协同相关物联实体的唯一标识和物联信息的统一调度。



（2）信息协同的模式和总体架构：根据智慧城市的大数据特征和需求，选择不同业务需求下的信息协同模式，通过改变信息的流向对传统模式下的信息流转进行优化，从体系、功能、数据和接口层面建立能够灵活适应多种信息协同模式的总体架构；根据信息协同不同阶段的特点建立信息流转的自适应进程，实现信息协同技术流程与业务流程的分离，最终实现信息在城市系统中智慧地流转。

（3）信息协同网络的结构测度和优化策略：对城市系统下的多源信息协同网络结构进行科学、有效的测度，包括信息组织的差异测度和信息组织间的关系测度，并从横向和纵向两个层面进行协同模式优化。

智慧城市多源信息协同体系的研究处在城市化和信息化的交叉领域，涉及城市管理、信息科学、计算科学、系统科学、决策科学、模糊数学、统计学、社会网络分析等多个学科，是一个非常复杂的系统工程。

3.3.3 智慧城市多源信息协同体系的理论与实践意义

1. 从信息视角解读智慧城市运行管理的内涵和机理

智慧城市作为一个开放的复杂巨系统，信息的协同是保障其他资源要素优化配置的基础，是城市系统更加智慧运行的前提。当前，智慧城市建设在解决一个个信息孤岛的同时，不可避免地又形成了领域间的新的智慧孤岛。从信息协同的角度重新解构智慧城市运行管理面临的问题和挑战，通过信息的流转分析不同事件驱动和不同场景边界下的城市运行资源调度，对诠释智慧城市的内涵和机理具有重要的理论意义。

2. 为城市和区域层面的信息协同体系建设提供理论依据和方法

1）为城市运行管理领域的信息协同标准提供理论依据和方法

智慧城市运行管理的信息协同标准体系是对智慧城市标准体系的扩展和细化，为城市运行领域和信息协同领域的系列标准建设提供了内部结构和外部接口上的参考。信息协同的关键标准为相关领域技术标准的制定提供了理论依据和方法。

2）为智慧城市的信息协同水平测度提供方法论的支撑

智慧城市多源信息协同的结构测度可以弥补总体测度的不足，为分析城市信息协同网络的内在结构及其关系提供合理、有效的方法和路径；同时，在差异测度和关系测度基础上的相应优化策略是信息协同模式自适应过程的理论基础。

3. 为大中型智慧城市建设的信息化顶层设计提供支撑

以跨部门、跨领域、跨层级的信息共享、流转和整合为基础的信息协同模式和通用体系架构可以应用于地方大中型智慧城市规划和建设中，从协同模式、协同机制和技术架构三个层面上指导智慧城市的信息化顶层设计，从信息流的角度为全面掌控城市系统运行情况、提高城市精细化管理水平和应急决策水平提供有效支撑。



第4章　智慧城市多源信息协同的标准体系

城市系统是一个开放的复杂巨系统。在后面的三章中，将以智慧城市为对象，分别阐述多源信息协同的标准体系、协同模式和总体架构。

信息协同标准是多源信息协同体系的基础，对信息协同的模式和评价进行规范性约束。本章重点介绍信息协同的标准体系框架及其关键标准，通过标准体系将智慧城市所涉及的人、地、事、物、组织用信息关联起来，最终实现智慧城市系统中所有实体的唯一标识和相关信息的统一调度。

4.1　智慧城市标准化现状

4.1.1　主要的信息化标准化组织



1. 国际信息化标准化组织

国际上的信息化标准化组织主要有国际标准化组织(ISO)、国际电工委员会(IEC)、国际物品编码协会(GSI)、美国电气和电子工程师协会(IEEE)、国际电报联盟(ITU)、欧洲电信标准化协会(ETSI)等。

1) ISO/IEC

国际标准化组织(International Organization for Standardization，ISO)成立于1947年2月23日，是世界上国际标准最大的推动者。中国加入ISO的时间是1978年，并在2008年10月的第31届国际化标准组织大会上正式成为ISO的常任理事国。代表中国参加ISO的国家机构是国家标准化管理委员会。

国际电工委员会(International Electrotechnical Commission，IEC)是国际性电工标准化机构，负责有关电气工程和电子工程领域中的国际标准化工作。IEC与ISO有密切的联系，ISO和IEC作为一个整体担负着制订全球

协商一致的国际标准的任务。代表中国参加 IEC 的国家机构是国家标准化管理委员会。

ISO/lEC 开展的相关信息化标准化工作有：针对 RFID、智能传感器等物联网相关关键领域开展了标准化工作；ISO 成立了智能电网特殊组；IEC 制定了智能传感器标准体系；ISO/IEC 成立了联合技术委员会(JTC)，针对标识、RF 接口、数据采集等开展工作；ISO/IEC JTC 成立了特殊组 WG7，专门研究传感器网络相关标准。

2）GSI

GS1（Globe Standard 1）即国际物品编码协会，其前身为 EAN International，成立于 1977 年，是基于比利时法律规定建立的一个非营利性质的国际组织，总部设在比利时首都布鲁塞尔。GSI 在 108 个国家设有办事处，有超过 2000 个专业人员。

2003 年 11 月 1 日，国际物品编码协会正式接管了 EPC（Electronic Product Code，产品电子代码）在全球的推广应用工作，成立了 EPC global，负责管理和实施全球的 EPC 工作。在我国，EPC global 授权中国物品编码中心作为唯一代表负责我国 EPC 系统的注册管理、维护及推广应用工作。同时，EPC global 于 2003 年 11 月 1 日将 Auto-ID 中心更名为 Auto-ID Lab，为 EPC global 提供技术支持。

此外 EPC global 还负责：参与 EPC 商业应用案例实施和 EPC global 网络标准的制订；参与 EPC global 网络、网络组成、研究开发和软件系统等的规范制订和实施；引领 EPC 研究方向；认证和测试；与其他用户共同进行试点和测试。

3）IEEE

美国电气和电子工程师协会（Institute of Electrical and Electronics Engineers，IEEE）在 1963 年 1 月 1 日由美国无线电工程师协会（IRE）和美国电气工程师协会（AIEE）合并而成，总部在美国纽约市。IEEE 拥有 300 多个地方分会，这些分会分布在 150 多个国家。



IEEE被国际标准化组织授权为可以制定标准的组织，设有专门的标准工作委员会，参与标准的研究和制定的工作人员达30 000多人，每年制定和修订800多个技术标准。IEEE定义的标准在工业界有极大的影响。其中802委员会于1980年2月成立，在制定局域网的国际标准方面成绩显著。

IEEE开展的相关信息化标准化工作有：IEEE 802系列标准是IEEE 802 LAN/MAN标准委员会制订的局域网、城域网技术标准，其中的IEEE 802.15工作组专门从事无线个人局域网（WPAN）标准化工作，IEEE的802.15工作组也是目前物联网领域在无线传感网层面的主要标准化组织之一。目前，传感器应用最广泛的ZigBee采用IEEE 802.15.4标准作为物理层和媒体存取控制层的标准。



4）ITU

国际电信联盟（International Telegraph Union，ITU）是联合国的一个专门机构，其总部在日内瓦。它由电信标准化部门（ITU-T）、无线通信部门（ITU-R）和电信发展部门（ITU-D）组成。其中，电信标准化部门由原来的CCITT和从事标准化工作的部门CCIR合并而成，主要职责是研究电信技术、操作和资费等问题，并出版了建议书，目的是在世界范围内实现电信标准化。

ITU开展的相关信息化标准化工作有：ITU-TSG 11主要研究标签和USN测试；ITU-TSG13主要研究NGN架构对RFID标签应用的支持，以及泛在网和泛在传感器网络需求和架构；ITU-TSG16主要研究标签和USN、UN业务相关工作；ITU-TSG 17开展针对标签、泛在网和泛在传感器网络安全方面的工作；此外还有智能电网、智能交通和IoT焦点工作组。

5）ETSI

欧洲电信标准化协会（European Telecommunications Standards Institute，ETSI）是一个非营利性的电信标准化组织，在1988年由欧盟批准建立，总部设在法国尼斯。ETSI的标准化领域主要是电信业，另外对信息及广播技术领域也有所涉及，但主要是与其他组织合作。ETSI获得了CEN

（欧洲标准化协会）和 CEPT（欧洲邮电主管部门会议）的认可，欧盟常把 ETSI 制定的推荐性标准作为欧洲法规的技术基础而采用并要求执行。

相比 ITU，ETSI 具有很大的公众性和开放性，主管部门、用户、运营商、研究机构都可以平等地发表意见。另外，它与 ITU 的不同之处还在于 ETSI 对市场敏感，根据市场和用户的需求制定标准，针对性和时效性较强。ITU 在制定标准时，常留有许多任选项给不同国家和地区进行选择，其结果是不便于设备的统一和互通。ETSI 针对欧洲市场和世界市场的情况，深入细化指标，避免了上述问题。

ETSI 开展的相关信息化标准化工作有：ETSI 在 2008 年 11 月成立了 M2M 技术委员会（M2M Technical Committee，M2M TC），其主要职责是收集和定义 M2M 需求、架构，补充现有标准没有覆盖的 M2M 的需求，并对这些需求进行标准化。M2M TC 的主要工作包括 M2M 设备标识、名址体系、QoS、安全隐私、计费、管理、应用接口、硬件接口、互操作等。

6）3GPP

第三代合作伙伴计划（The 3rd Generation Partnership Project，3GPP）标准组织创建于 1998 年 12 月。3GPP 的组织伙伴包括欧洲的 ETSI、日本的 ARIB、日本的 TTC、韩国的 TTA、美国的 TI 和中国通信标准化协会六个标准化组织。3GPP 致力于 3G 及长期演进分组域网络的研究。

3GPP 在这方面的工作包括：GERAN（GSMEDGE Radio Access Network）研究通过增强和优化 GERAN 支持 M2M 通信；RAN（Residential Access Network）研究通过增强和优化 UTRAN/E-UTRAN 支持 M2M 通信；SA1 负责确定 M2M 通信的基本需求、业务需求和 MTC Feature；SA2 负责研究 M2M 通信网络架构及网络优化；SA3 负责研究 M2M 通信中的安全问题。

2. 国内信息化标准化组织

国内制定信息化标准的机构主要包括中国通信标准化协会（CCSA）、国

家标准委员会传感器网络标准工作组(WGSN)、工业和信息化部电子标签(RFID)标准工作组以及各行业标准化组织。

1) 中国通信标准化协会(CCSA)

中国通信标准化协会(China Communications Standards Association,CCSA)于2002年12月18日成立。CCSA是国内企、事业单位自愿联合组织起来,经业务主管部门批准,国家社团登记管理机关登记,开展通信技术领域标准化活动的非营利性法人社会团体。

CCSA的主要任务是更好地开展通信标准研究工作,把通信运营企业、制造企业、研究单位、大学等关心标准化的企事业单位组织起来,按照公平、公正、公开的原则制订标准,进行标准的协调、把关,把高技术、高水平、高质量的标准推荐给政府,把具有中国自主知识产权的标准推向世界,支撑中国的通信产业,为世界通信做出贡献。



CCSA开展的物联网相关工作包括:TC3开展了泛在网的需求和架构、M2M业务相关的标准工作;TC5开展了WSN与电信网结合的总体技术要求、TD网关设备要求相关的标准工作;TC5开展了机器类通信安全相关的标准工作;TC10泛在网技术工作委员会,包括总体工作组、应用工作组、网络工作组、感知延伸工作组,专门研究泛在网相关的标准工作。

2) 传感器网络标准工作组(WGSN)

传感器网络标准工作组(WGSN)是从事传感器网络标准化工作的全国性技术组织,成立于2000年9月11日,由国家标准化管理委员会批准筹建,由全国信息、技术标准化技术委员会批准成立并领导。

传感器网络标准工作组的主要任务是根据国家标准化工作的方针政策,研究并提出有关传感网标准化工作方针、政策和技术措施的建议;按照国家标准制定、修订原则,制订和完善传感网的标准体系表。工作组提出和修订传感网国家标准的长远规划和年度计划的建议;根据批准的计划,组织传感网国家标准的制定、修订工作及其他与标准化有关的工作。

3）RFID 标准工作组

RFID 标准工作组在原信息产业部科技司领导下开展工作，专门致力于中国 RFID 领域的技术研究和标准制订，信息产业部电子标签标准工作组成立于 2005 年 10 月，下设 7 个工作组，包括总体组（47 家）、标签与读写器组（48 家）、频率与通信组（26 家）、数据格式组（14 家）、应用组（52 家）、信息安全组（18 家）、知识产权组（4 家）。目前我国正在研究和制定的标准超过 40 项。

电子标签标准工作组提出了中国的 RFID 标准体系（见图 4.1），并有针对性地开展了研究，我国 RFID 市场规模已居全球第三位。

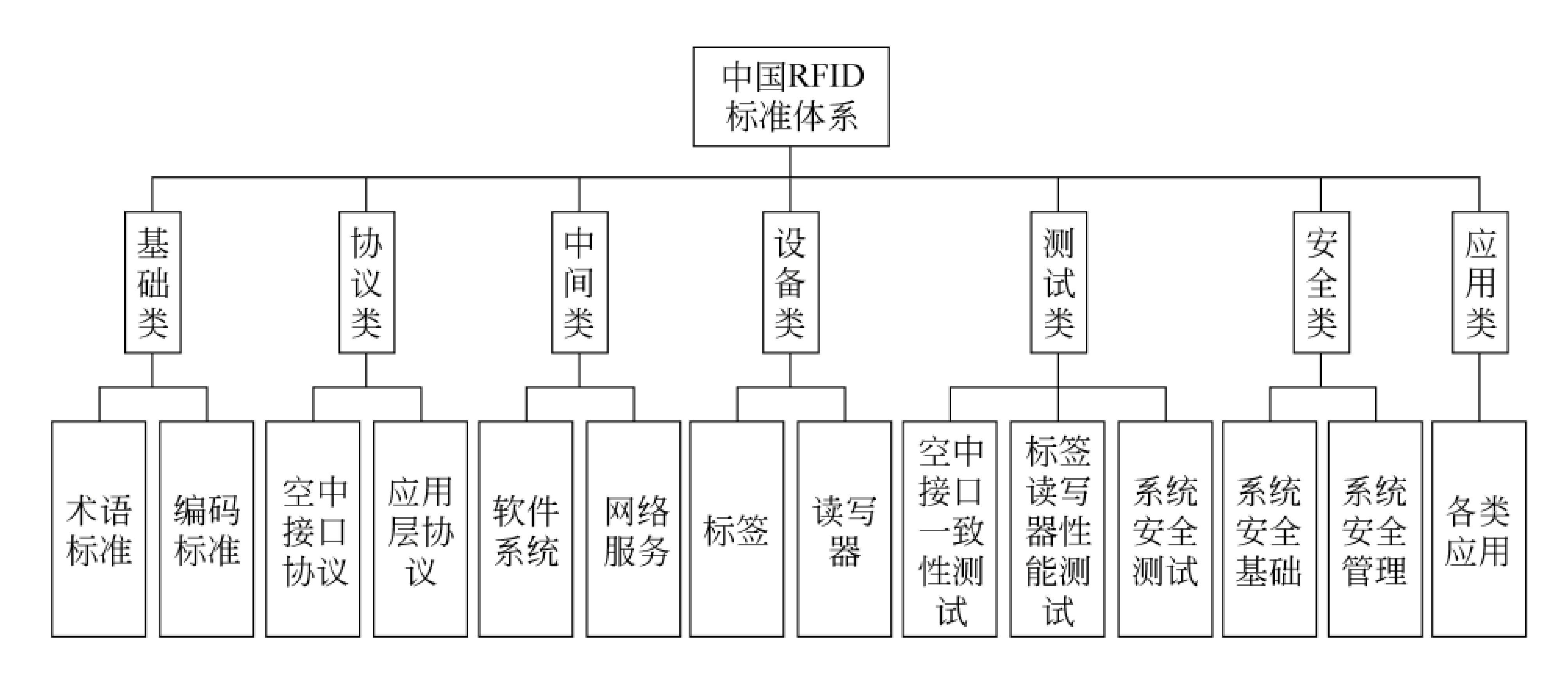


图 4.1 中国 RFID 标准体系

电子标签标准工作组成员单位参与制定的 RFID 标准主要有《GB 18937—2003 全国产品与服务统一标识代码编制规则》、《TB/T 3070—2002 铁路机车车辆自动识别设备技术条件》以及在上海市使用的《送检动物电子标示通用技术规范》。

RFID 标准工作组目前已经公布的相关 RFID 标准主要有参照 ISO/IEC 15693 标准的识别卡和无触点的集成电路卡标准，即《GB/T 22351，1—2008 识别卡无触点的集成电路卡邻近式卡第 1 部分：物理特性》和《GB/T 22351，3—2008 识别卡无触点的集成电路卡邻近式卡第 3 部分：防冲突和

传输协议》。

4）中国电子技术标准化研究所

中国电子技术标准化研究所(原信息产业部电子工业标准化研究所、信息产业部电子第四研究所，简称 CESI)是国际电子技术标准化权威研究机构。

该所长期参与国际标准化活动，承担了多项重要的安全和电磁兼容国家标准(如 GB9254、GB4943、GB/T17618、GB/T6113 系列)和国家军用标准(如 GJB151/152、GJB151A/152A)的制定和修订工作，对相关标准条款有最终解释权，对标准的理解和使用具有特殊的优势。



4.1.2 全球智慧城市标准化现状

ISO、IEC、ITU-T 均在积极推动智慧城市标准化工作。2012 年 2 月 23 日，ISO 响应联合国、世界银行等国际组织以及世界各国对可持续发展标准化的需求，批准成立了 ISO/TC268 社区可持续发展技术委员会(Sustainable Development in Communities)，目的是为推动世界各国城市(社区)实现可持续发展，为各类城市(社区)提供支撑技术和工具，包括管理体系要求、指南和相关标准。ISO/TC268 围绕城市和社区可持续发展，组建了 ISO/TC268SC1 城市智能基础设施计量分技术委员会(Smart Urban Infrastructure Metrics)，负责制定城市(社区)基础设施计量的标准。2011 年日本向 ISO TMB 提出了衡量城市(社区)基础设施智能程度的评估方法，对城市智能基础设施的定义、范围、计量方法等内容提出了具体要求，目的是通过定量分析的方法衡量城市能源、水、交通及 ICT 等城市基础设施，目前正在围绕城市(社区)基础设施标准的范围、术语、定义及通则等方面制订建议草案。在 2013 年 2 月，ITU-T 建立了一个新的针对智慧城市可持续发展问题的专题评估小组，用以评估智慧城市标准化工作。2013 年 6 月，IECSMB 同意设立智慧城市的系统评价小组，2013 年 7 月 9 日在日本召开了第一次工作会议。

欧美各国都在积极开展国家或区域智慧城市标准化工作。CEN/CENLEC(Comité Européen de Normalization，欧洲标准委员会/European Committee for Electrotechnical Standardization，欧洲电工标准化委员会)、BSI(Britain Standards Institute，英国标准协会)、DIN(Deutsches Institut für Normung，德国标准化协会)、ANSI(American National Standards Institute，美国国家标准学会)等地区或国家标准化机构纷纷提出了智慧城市标准化战略定位、体系框架和参考模型。2012年，BSI提出了智慧城市标准化战略，目前正在推动《智慧城市框架：智慧城市和社区决策者的良好做法指南》和《智慧城市术语》等标准的研制工作。2013年4月，CEN及CENELEC共同成立了可持续的智慧城市和社区协调小组(SSCCCG)，加速推进欧洲的智能城市标准化进展。2013年4月4日，ANSI召开了联合会员论坛，讨论标准和一致性解决方案在促进国家和国际智慧城市建设过程中发挥的重要作用。2013年5月，DIN和DKE(德国电气电工信息技术委员会)发布了一份题为"德国标准化路线图：智能城市和社区的可持续发展"的报告。

亚洲国家也纷纷针对智慧城市提出了各国的标准研究计划。2006年，新加坡推出了智慧国2015计划，并开始研究智慧城市标准化工作。2007年，韩国成立了U-Eco City的研发机构，其任务之一是研究智慧城市的标准化工作。2011年，日本INOTEK组织也开始研究智慧城市标准化工作。从2011年开始，我国国家标准化管理委员会(SAC)也在积极开展智慧城市标准化工作。

4.1.3 国内智慧城市标准化现状

智慧城市建设的标准化有利于提高城市规划的通用性，降低建设成本，有利于促进产业链的分工合作，加强各政府部门之间的互联互通、共享协同，推广最佳的技术和应用实践。

国内智慧城市标准体系的研究和关键标准的制定工作与国外处于同步

发展阶段，主要的标准化机构和组织包括国家标准化管理委员会(Standardization Administration of China，SAC)、全国通信标准化技术委员会、全国信息技术标准化技术委员会(TC28)、全国智能建筑及居住区数字化标准化技术委员会(TC426)、全国智能运输系统标准化技术委员会(TC268)、中国通信标准化协会(CCSA，TC10)、全国信息分类与编码标准化技术委员会(TC353)等。

2013 年 10 月，SAC 联合科技部开展智慧城市标准的试点、验证和示范工作。2014 年 1 月，SAC 联合 9 个部委启动智慧城市国家标准体系建设工作，首批下达了《智慧城市技术参考模型》等五项国家标准，以指导推进智慧城市国家标准制定、国际标准化工作以及标准服务体系建设。《智慧城市技术参考模型》适用于智慧城市整体规划及具体领域信息化项目的设计、开发、运行和维护，是指导和综合应用智慧城市的具体技术、服务实现标准的依据，也是建立智慧城市相关质量测评标准、工程标准及应用标准的依据。

2014 年 7 月，全国首批 17 个智慧城市专项联合实验室揭牌，聚焦城市基础设施、城市安全、水务工程、建筑节能、智能交通、公共信息平台、信息安全、信息技术等领域的科研和标准建设。

同时，住房和城乡建设部编制了《智慧城市评价模型及基础评价指标体系》的国家标准。《智慧城市评价模型及基础评价第 1 部分：信息基础设施》由全国通信标准化技术委员会牵头负责，规定智慧城市信息基础设施评价对象、范围和指标，并提出相应的评价指标。其中评价指标中包括技术原则和要求以及设计与使用原则。本标准适用于智慧城市整体规划及信息基础设施项目建设与评价，是指导智慧城市具体技术、服务实现的标准依据，也是建立智慧城市相关质量测评标准、工程标准及应用标准的依据。《智慧城市评价模型及基础评价第 2 部分：信息化应用和服务》由全国信息技术标准化技术委员会牵头负责，规定智慧城市信息化应用与服务评价模型、评价指标，适用于智慧城市整体规划及信息化应用与服务项目建设与评价，是评估后续智慧城市具体应用与服务标准的依据。《智慧城市评价模型及基础评

价指标体系第 3 部分：建设管理》由全国智能建筑及居住区数字化标准化技术委员会牵头负责，标准中建设管理内容主要指城市建设中的水、电、煤气等基础设施管网的建设标准，结合移动互联网、物联网、云计算等先进信息技术与城市管理运营理念，致力于提高城市的基础设施的协同化、智慧化，提供城市生活的舒适度。

除此之外，全国信息技术标准化技术委员会牵头负责的《智慧城市 SOA 标准应用指南》规定了智慧城市的 SOA 应用参考模型及智慧城市建设中 SOA 标准的综合应用建议，适用于智慧城市整体及具体领域信息化项目的规划、设计、开发、实施、评估、运行和维护；全国信息安全标准化技术委员会负责的《信息安全技术智慧城市建设信息安全保障指南》针对智慧城市建设带来的数据资源集中和共享所面临的信息安全风险，进行信息安全保障体系的标准制定。

在地方城市层面，部分省市已开展了智慧城市标准的研究和先期应用。浙江省、上海市、南京市等地方已将智慧城市标准工作纳入工作任务，并成立了地方标准化组织，积极开展智慧城市评价指标体系、体系结构、信息资源目录和交换等标准规范的研究工作。

根据相关公开的计划和进展，到 2016 年年底国内将逐步建立健全中国智慧城市国家标准体系。其中，共性、关键性标准主要涵盖数据与服务融合平台、主数据、数据挖掘分析、跨系统信息交互、信息资源管理与信息系统运维等促进系统、数据与服务融合的领域。

4.2 信息协同标准化的关键问题与总体框架

4.2.1 城市化、信息化与标准化协同发展

2014 年 3 月 16 日发布的《国家新型城镇化规划（2014—2020 年）》明确提出了推进智慧城市建设，统筹城市发展的物质资源、信息资源和智力资源

利用，推动物联网、云计算、大数据等新一代信息技术创新应用，实现与城市经济社会发展深度融合的理念。作为城市发展的高级阶段，智慧城市成为信息化与城市化高度融合的最佳契合点，也是推进城市化建设的战略制高点。

当前，我国的智慧城市建设处于基础设施建设和领域示范应用的起步阶段，智慧城市的架构模式、标准规范、关键技术、评价体系等均不成熟。统计数据表明，目前全国所有副省级以上城市、超过 89％的地级及以上城市、超过 47％的县级及以上城市提出了智慧城市建设的设想或规划。但是，在智慧城市的实际建设过程中，普遍面临着重应用轻规划、重建设轻标准、重技术轻规范的问题。从顶层设计的层面上构建完整、合理的标准体系框架，已成为指导智慧城市科学建设和可持续发展的关键。

根据钱学森、于景元、戴汝为在《一个科学新领域——开放的复杂巨系统及其方法论》中的定义，智慧城市是一个开放的复杂巨系统，城市系统本身与系统周围的环境有物质、能量和信息的交换，同时城市系统下又包含数量庞大、种类繁多的子系统。智慧城市的标准化建设是促进城市各利益相关方达成共识的基础，同时也是促进城市产业链优化、降低城市运行成本、提升城市竞争力的重要保障；而信息协同则是保障城市系统中其他资源要素优化配置的基础，是城市系统更加智慧运行的前提。智慧城市建设需要城市化、信息化和标准化的协同发展（见图 4.2）。

物联网、云计算、移动互联网、大数据等新一代信息技术的应用带来技术变革的深入发展，技术标准体系建设成为标准壁垒时代技术创新的重要支撑。随着信息系统的泛在化逐步成为全球信息化向高端发展的主要特征，城市管理需要城市各部门、各领域、各主体间信息的互联互通，信息共享的重要性日益凸显。但是，智慧城市的建设过程在解决一个个“信息孤岛”的同时，不可避免地又形成了领域间新的“智慧孤岛”，各领域应用均按照各自的管理思路和标准体系建设，在城市和区域层面上缺少统筹协调。智慧城市管理的信息协同标准体系成为城市化、信息化和标准化协同发展亟待

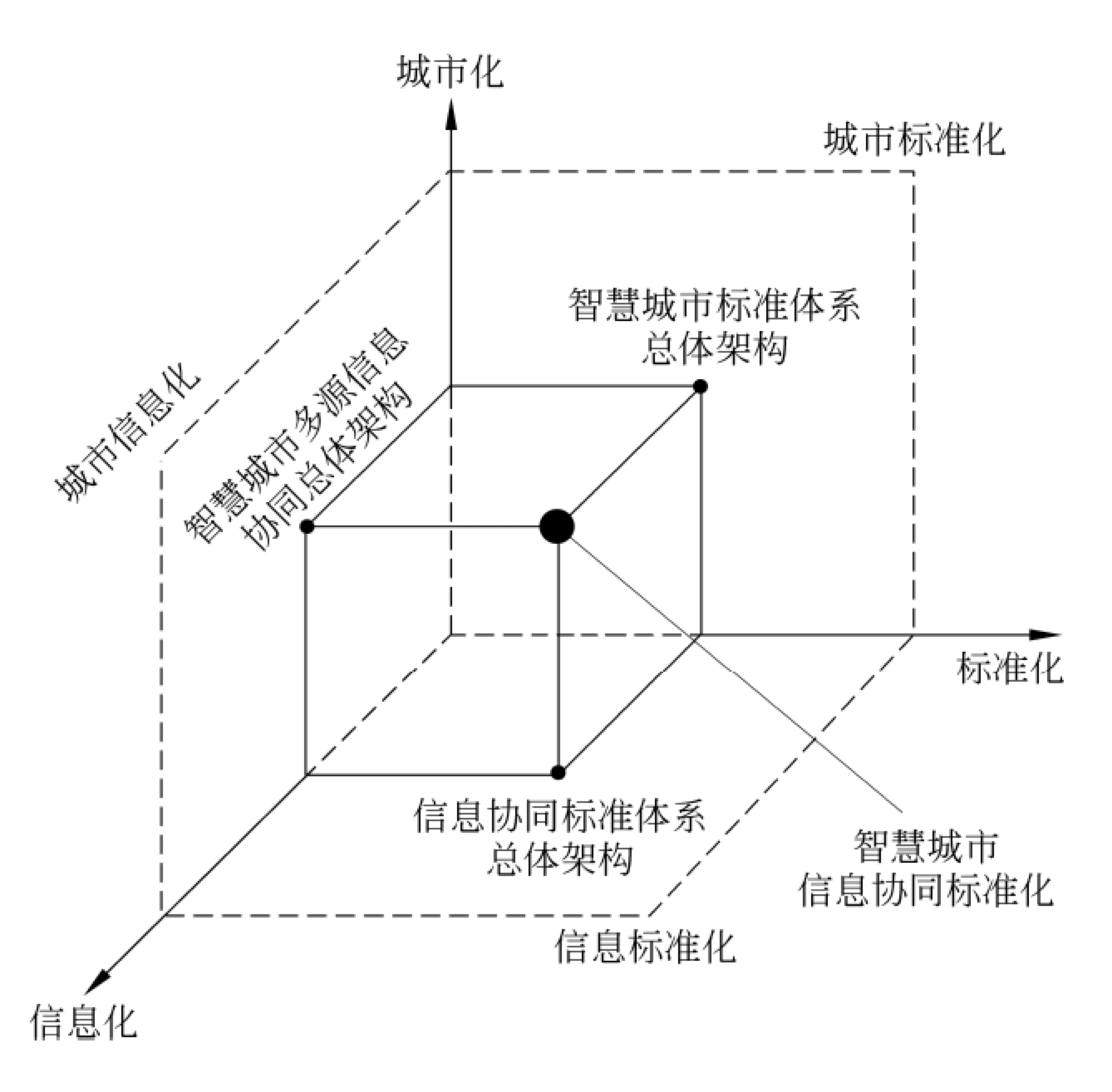


图 4.2　城市化、信息化与标准化协同发展

解决的核心问题。

4.2.2　智慧城市信息协同标准的关键问题

随着新一代信息技术的发展和全球智慧城市的规模化建设，感知设备越来越多地应用在城市管理的各个层面。通过感知设备对管理对象的实时感知，能够为城市运行的监管和服务提供更快速、更有效的数据支撑。面对日益增多、分布在城市各个角落的大量传感设备，如何对其进行科学的管理，同时加强感知设备和感知信息的共享和利用，保障感知信息在城市系统下合理、高效的流转，是智慧城市信息协同标准体系需要解决的关键问题。

第一，物联实体的编码是智慧城市标准化建设和应用的基础。一方面，标准的编码及其表示方式有利于感知设备、管理对象等城市运行管理中物联实体的唯一识别；另一方面，实时感知信息是物联信息的主体部分，通过编码可以将不同时间采集的信息进行统一管理，实现物联实体的实时定位。

第二，统一的基础信息属性管理是城市基础设施集约化建设的前提。



一方面,属性规范有利于物联实体管理的规范化和标准化,确立编码与信息之间的一一对应关系,从而保证对信息表述的唯一性、可靠性和可比性;另一方面,统一的基础信息属性管理有利于城市基础设施的重复利用和有效利旧,可以大幅减少政府和市场的投资。

第三,物联信息在城市系统下的流转需要统一模式规范和技术标准,并与物联实体的基础信息属性和编码标准相关联。一方面,根据实体编码规则对采集信息进行不同维度的分类,形成有效的信息,便于在信息流转过程中进行整合融合;另一方面,信息流转过程中包含城市系统下的唯一编码,并通过基础属性描述与物联实体相关联,有利于根据不同的业务和场景需求选择合理的信息流转模式,并根据需求的变化进行自适应调整。

4.2.3 智慧城市多源信息协同的标准体系总体框架

目前,国内外在智慧城市标准体系框架和共性标准方面尚未形成统一的认知。一般认为,智慧城市标准体系涉及总体标准、基础设施、建设与宜居、管理与服务、产业与经济、安全与运维等多个部分。智慧城市多源信息协同标准作为智慧城市标准体系框架的主要组成部分,需要包含信息技术标准体系和城市管理标准体系两个范畴的必要元素;同时,还应体现智慧城市不同于传统城市和数字城市的典型特征。智慧城市多源信息协同标准体系的总体框架如图 4.3 所示。



从内部结构上看,智慧城市多源信息协同标准体系主要由五个子体系组成:

(1) 基础信息标准体系:重点在于物联实体的基础信息属性、编码及其表示方式。

(2) 信息流转标准体系:重点在于城市系统下的信息共享方式、信息流转模式及其共性规范。

(3) 领域应用标准体系:以业务属性为主,具体内容由行业的要求和特点决定。

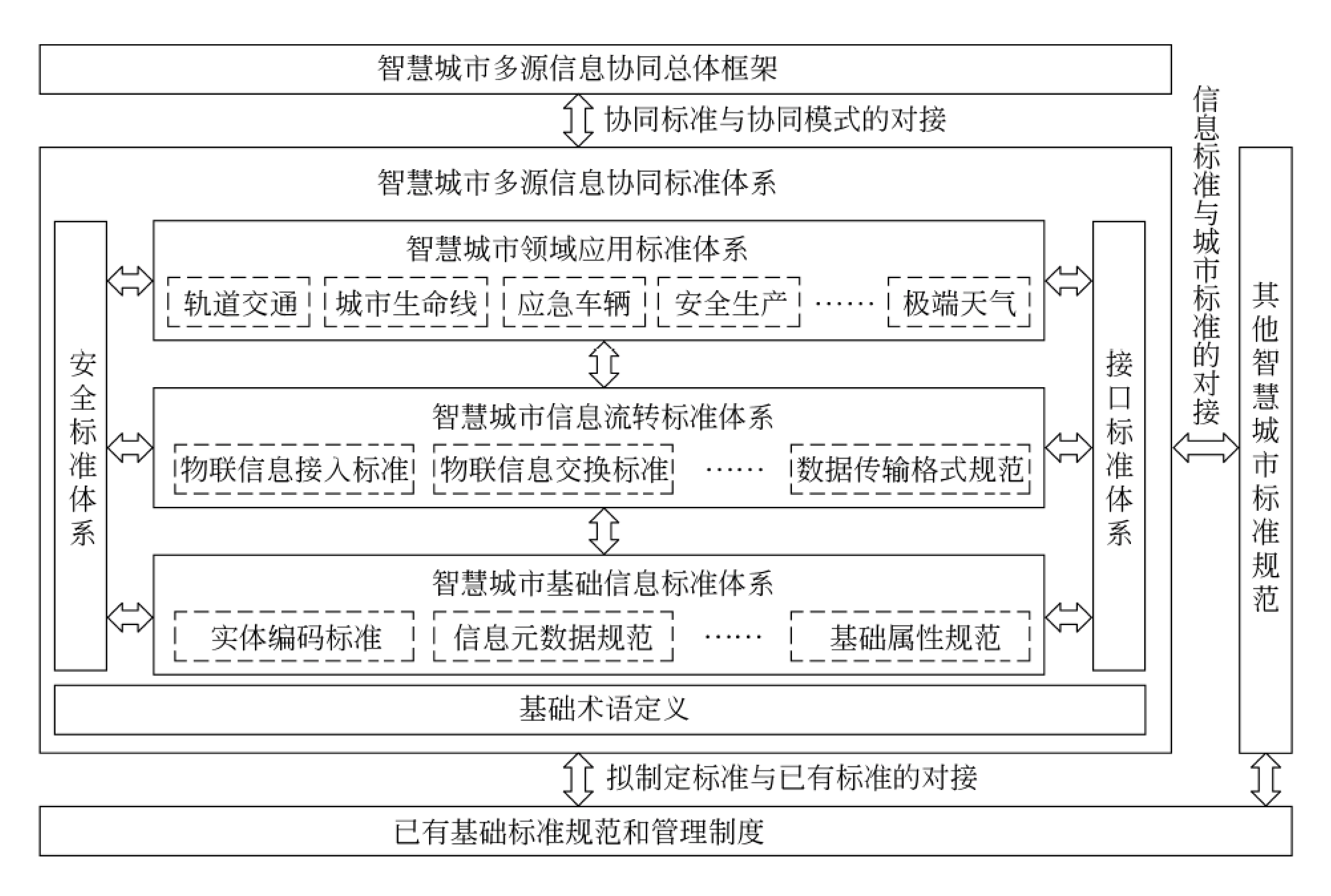


图 4.3　智慧城市多源信息协同标准体系的总体框架

(4) 安全标准体系：主要包括安全防护体系的基本规范，重点涉及身份验证、访问控制、传输安全等技术要求。

(5) 接口标准体系：重点在于信息流转过程中数据交换接口规范，及智慧城市系统中与识别和定位相关的认证授权接口规范。

从内部关系上，基础信息标准、信息流转标准和领域应用标准在逻辑关系上自下而上，下层标准是上层标准的基础；安全标准和接口标准属于标准体系的共性规范，与其他所有标准相关联。

除此之外，智慧城市多源信息协同标准体系需要做好三个层面的外部对接：第一，与智慧城市多源信息协同总体框架之间，形成标准体系与协同体系对接；第二，与智慧城市标准体系中的其他标准规范之间，形成信息标准与城市标准的对接；第三，与已有基础标准规范和管理制度之间，形成规划标准与已有标准的对接。

智慧城市多源信息协同标准体系建设是打通领域间的“智慧孤岛”、实



现城市和区域物联实体"一盘棋"管理的必经之路，也是标准化与城市化、信息化协同发展亟待解决的核心问题。通过对智慧城市系统中的各类实体和多源信息的统筹管理，能够为智慧城市的运行监管和精细化服务提供更快速、更有效的支撑。

智慧城市中的各相关部门和各领域主体应在信息协同标准体系总体框架基础上，分别制定本行业、本领域的应用和服务标准，实现技术标准与服务标准、信息标准与城市标准、规划标准与现行标准的一体化管理；同时，在标准体系的建设过程中充分体现信息技术和城市管理两个范畴的核心内涵，做好标准体系与信息协同总体框架的无缝对接，体现智慧城市不同于传统城市和数字城市的典型特征，逐步建立完善的智慧城市信息协同标准体系。

4.3 智慧城市多源信息协同的关键基础标准



4.3.1 基础信息属性、编码及其表示方式

基础信息标准系列主要以智慧城市管理的应用需求为依托，以城市系统下的物联信息流转为前提，对管理对象、感知设备、传感网网关等物联实体的基础信息特征所需要的属性、编码及其表示方式进行规定，用于不同应用领域对物联实体基础信息的统一标识和统筹管理。

1. 基础信息属性

物联实体的基础信息可以按照管理要求分为核心属性和扩展属性。其中，核心属性属于强制管理范畴，扩展属性主要体现业务管理特性，可根据物联实体的个性特征进行扩展。

1）管理对象基础信息

管理对象基础信息的核心属性主要包括管理对象名称、管理对象编码（唯一管理标识编码）、管理对象其他编码（在其他编码体系或应用系统中存在的编码）、管理对象类别码（所属分类）、管理对象所属业务、管理对象空间

位置(包括位置描述,坐标类型,坐标单位,横、纵坐标等)、管理对象负责方(包括负责部门、重点监控和管控的内容等,如危险品仓库重点关注存储环境安全、温湿度等信息)。

管理对象基础信息的扩展属性主要体现管理对象具体的业务管理特性。如“烟花爆竹仓库”具有“库房间数、仓库面积、灭火器数、消防水源、限制存箱量、限制存药量”,“烟花爆竹批发单位”具有“安全管理人员数量、仓库保管和守护人员数量、运输车辆数量”,“消防车辆”具有“车辆牌号”,“自动气象站”具有“建设时间”等特有属性。

2)感知设备基础信息

感知设备基础信息的核心属性主要包括感知设备名称(如温度传感器)、感知设备编码(唯一管理标识编码)、感知设备其他编码(在其他编码体系或应用系统中存在的编码)、感知设备类别码(所属分类)、是否固定(感知设备工作时是否固定安装在某个位置)、感知管理对象编码(感知设备用来感知的管理对象的编码,感知设备与管理对象的关系为多对一的关系)、设备用途(感知设备用来感知的内容描述)、感知频率(感知设备采集信息的最小时间间隔)、编解码标准(感知设备在传输过程中采用的信息编解码标准,包括编解码标准名称(如 SVAC 标准)、编解码标准类型(如 SVAC 通用标准)、编解码方式等)、感知设备空间位置(包括位置描述,坐标类型,坐标单位,横、纵坐标,高度等)、感知设备所属业务、感知设备负责方、感知设备型号、感知设备安装时间、感知设备生产时间、感知设备生产单位等。

感知设备基础信息的扩展属性主要体现的感知设备个性特征管理属性。如感知设备具有的参数信息,如安全监管部门的温度计具有“所属仓库、所属库房、所属房间、型号、最小刻度、最大刻度”,环保部门的空气质量传感器具有“量程、工作温度、存储温度、加热阻抗、加热电压”等特殊属性。

3)传感网网关基础信息

传感网网关基础信息属于强制管理范畴。核心属性主要包括传感网网关名称、传感网网关编码(唯一管理标识编码)、感知设备其他编码(在其他



编码体系或应用系统中存在的编码)、接入的感知设备编码、传感网网关连接的CPE(包括CPE的IP地址、SIM卡号(主要针对采用无线网传输的CPE)等)、传感网网关空间位置(包括位置描述,坐标类型,坐标单位,横、纵坐标,垂直高度等)、传感网网关所属业务、传感网网关负责方、传感网网关型号、传感网网关安装时间、传感网网关生产时间、传感网网关生产单位等。

管理对象、感知设备、传感网网关的基础信息核心属性及其关联关系如图4.4所示。

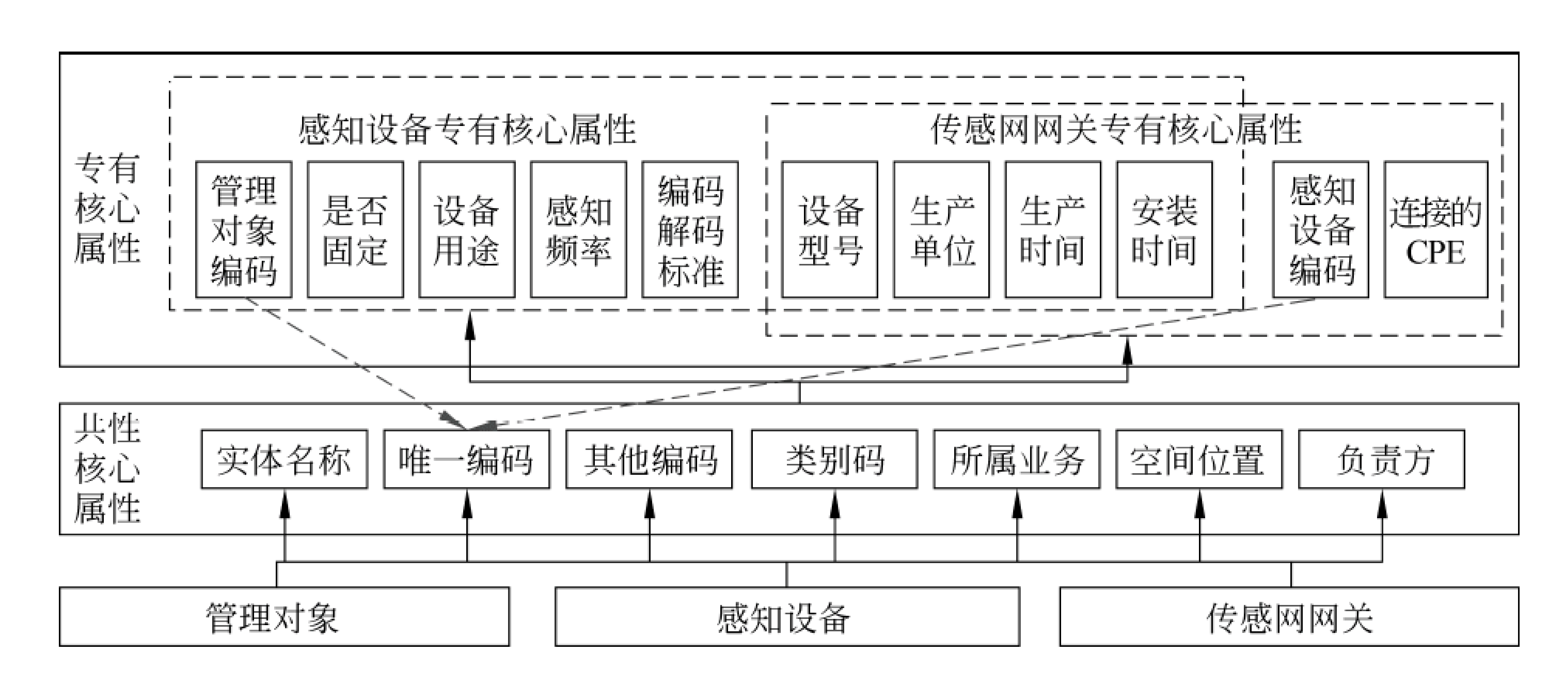


图4.4 物联实体基础信息的核心属性及其关联关系

感知设备基础信息的核心属性中,通过管理对象编码与对应感知的管理对象进行关联,感知设备与管理对象的关系为多对一关系;传感网网关基础信息的核心属性中,通过感知设备编码与所接入的感知设备进行关联。

2. 物联实体编码结构与管理

物联实体的编码规则需要体现三个特点:一是唯一性,即实体的唯一标识;二是可分类,即编码具有一定的规则;三是能够反映所编码的承载信息,即数据包含有意义的内容。

为了编码规范的通用性和使用的灵活性,编码结构应由定长码(前段码)和非定长码(后段码)两组编码组成。定长码(前段码)由编码管理部门统一分配和维护;非定长码(后段码)根据不同领域和行业的业务特征,由物

联实体的所属部门依据现有标准或自行编制，并将编码规则向编码管理部门提交备案。

1）定长码（前段码）

定长码（前段码）规定编码版本、实体类别、实体归属部门等基本信息，由编码管理部门统一管理。其中，版本码 2 个字节，由编码管理部门负责发布编码版本代号，根据编码中相关因素变更情况进行动态调整，并定期统一更新；类别码 1 个字节，“1”代表管理对象，“2”代表感知设备；单位编码使用物联实体所属部门的组织机构代码（或根据相关部门的统一标准规定执行），由编码管理部门负责发布和变更，变更情况在版本码中体现。

2）非定长码（后段码）

感知设备的后段码编码结构采用“感知设备出厂标准编码＋[顺序码]＋校验码”的形式。其中，感知设备出厂标准编码也可以采用现有国家或行业或地方标准编码或自定义码。如果可以保证唯一性，则顺序码可省略。

管理对象的后段码编码结构采用“现有标准分类编码＋[顺序码]＋校验码”的形式。兼容现有标准分类编码可以由多个标准编码组合，如可以由市政设施部件编码＋位置编码＋行政区划编码，也可以仅采用单一的标准编码。如果现有标准分类编码能够保证唯一性，则顺序码可省略。



3）码段补位

编码时，如采用十进制编码方式，长度一般不超过 90 位；如采用十六进制编码方式，长度一般不超过 0xF6（即十进制数 246）位。当编码长度小于最大长度时，不足位补零。如采用自定义码，且自定义码为顺序码时，待编码的管理对象或感知设备数量为 2000 个，编码起始号为 1，结束号为 2000。由于编码总长度不得小于 24 位，减去前段码长度 12 位后，后段码长度为 12 位，则后段码编码为 000000000001 至 000000002000。

非定长码（后段码）发生变更或部门内部编码规则发生变更时，应向编码管理部门提交申请并同步更新。此外，编码应保证整体结构的完整性，即定长码和非定长码在编制和使用时不得拆分。



4.3.2 物联信息的接入与交换接口

1. 物联信息接入

信息流转标准主要规范智慧城市中可共享信息的基本描述、流转模式与技术要求，实现信息在城市系统下共享和接入的统筹管理，并对信息在流转过程中的整合融合标准提供共性支撑。

从城市系统下的信息流向来看，信息的流转路径主要包括信息接入和信息协同两种情况（见图 4.5）。信息接入模式主要是信息的单向流转，包括直接流转、接入转发和服务接入三种情况；信息协同模式则主要以前置交换的形式开展跨部门、跨层级的信息联动。

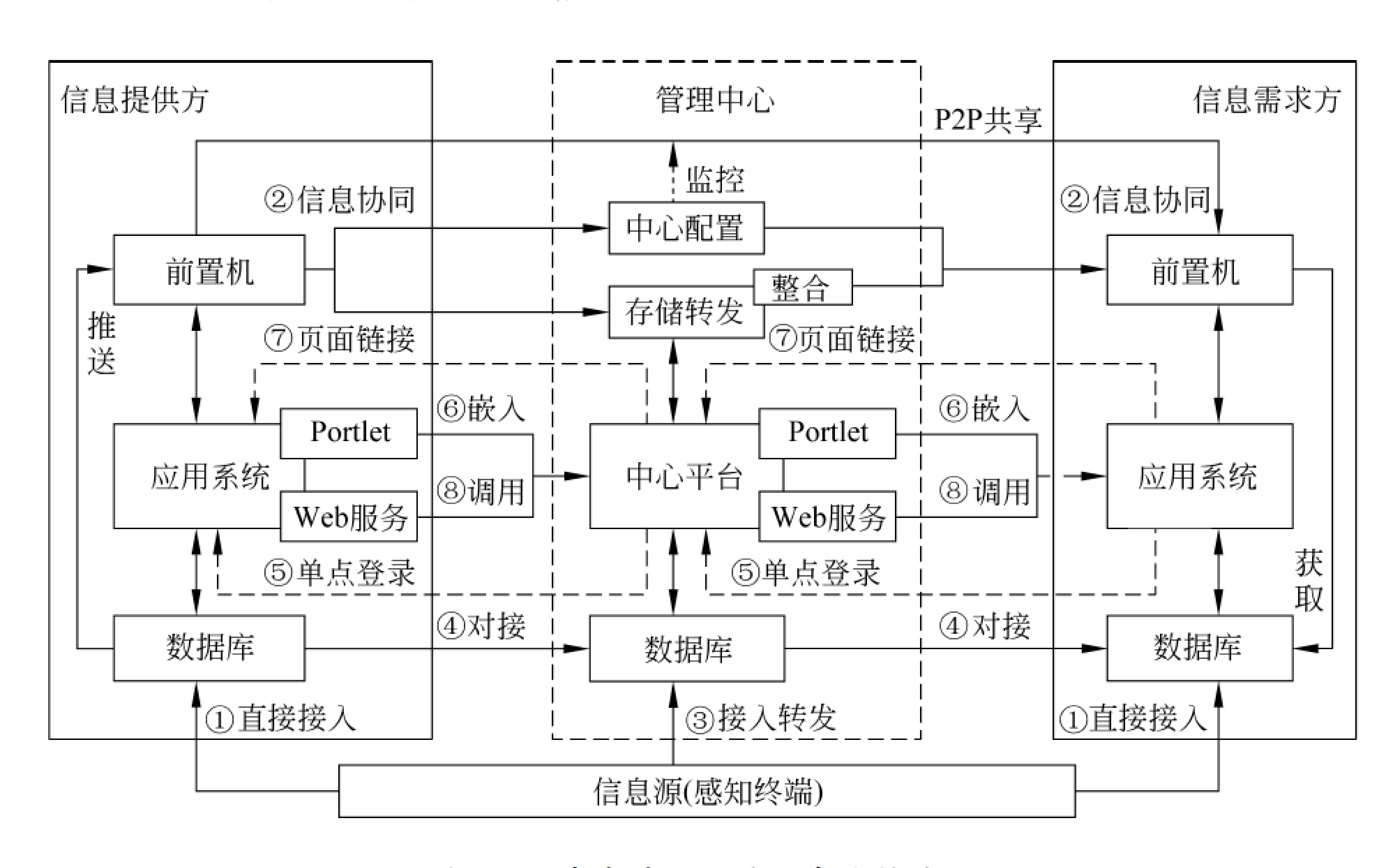


图 4.5　城市系统下的信息流转路径

1）信息接入模式

（1）直接流转方式：包括直接接入和数据库对接两种方式。

直接接入指信息从信息源直接接入需求方的数据库，主要对数据库类型、数据结构等基本信息描述和技术要求进行规范。

数据库对接指直接(或通过统一的中心平台数据库)访问信息提供方的数据库获取信息,主要对数据库类型、用户名、连接方式(如 JDBC 或 ODBC)等基本信息描述和技术要求进行规范。

(2) 接入转发方式:实时感知信息统一接入中心平台后,根据实际业务需要和约定规则直接分发至一个或多个信息需求方,主要对数据类型、数据需求方等基本信息描述和技术要求进行规范。

(3) 服务接入方式:包括系统对接、Portlet 嵌入、页面链接和 Web 服务接入四种方式。

系统对接指采用单点登录的方式,直接(或通过统一的中心平台)访问其他授权的应用系统,主要对系统名称、系统结构(如 B/S 结构)、单点登录地址、数据同步地址、同步数据类型、系统改造(接口要求)等基本信息描述和技术要求进行规范。

Portlet 嵌入指信息提供方提供 Portlet,直接(或通过统一的中心平台)嵌入信息需求方的应用系统,主要对页面名称、链接地址、编码要求(如 URL 编码)、通信协议(如 HTTP 协议和 HTTPS 协议)、代码规范(如符合 HTML4.0 规范或 XHTML1.0 规范)等基本信息描述和技术要求进行规范。

页面链接指信息需求方直接(或通过统一的中心平台)链接信息提供方的页面地址,主要对页面名称、宽度、高度、地址、登录验证、技术规范(如符合 JSR-168 规范)等基本信息描述和技术要求进行规范。

Web 服务接入指信息需求方直接(或通过统一的中心平台)调用信息提供方发布 Web 服务,主要对 Web 服务名称、地址、类别、方法、有效期、服务协议要求(服务发现协议、服务传输协议、服务消息协议)等基本信息描述和技术要求进行规范。

2) 信息协同模式

按照约定的交换规则,信息需求方通过前置机的方式获取信息提供方的信息,主要对数据类型、交换频率、转换规则等基本信息描述和技术要求



进行规范。

信息流转标准体系一方面需要对不同信息共享和接入方式的基本信息描述和技术要求进行规范，另一方面需要对数据传输格式等共性的技术标准进行统一。

2. 物联信息协同接口

物联信息协同的接口主要包括数据库协同接口和文件协同接口两种方式。每种方式提供 JMS(Java Message Service，Java 消息服务)和 Web 服务两种接口，并支持跨语言、跨操作系统调用。

1）数据库协同接口

数据库协同根据数据库数据交换请求，将特定的数据库数据交换到指定的其他信息协同节点，支持各种主流关系型数据库(包括 Oracle、SQL Server、DB2、Access、MySQL 等)之间的数据协同。

(1）数据库协同请求接口：将数据库协同请求信息发送到管理中心，并在发送数据的信息协同节点与接收数据的信息协同节点之间建立会话。数据库协同请求接口参数说明见表 4.1(数据库协同请求对象 DBExRequest 的 XML Schema 定义见附录 C)。

表 4.1　数据库协同请求接口参数说明

参数名称	参数类型	可选/必选	参数含义
Message	数据库协同请求对象 DBExRequest	必选	数据库协同请求信息
deliveryMode	Int	可选	采用的传送模式，支持的传送模式至少应包括持续传送模式 PERSISTENT 和一次性传送模式 NON_PERSISTENT
Priority	Int	可选	请求的优先级，共有 10 个优先级。0 是最低优先级，9 是最高优先级。默认的优先级是 4
timeToLive	Int	可选	请求的存在时间，单位为毫秒，由发送方指定，0 为无限制

（2）数据库数据发送/接收接口：将需交换的数据库数据发送给接收数据的信息协同节点。数据库数据发送/接收接口参数说明见表4.2（数据库协同数据对象DBExData的XML Schema定义见附录C）。

2）文件协同接口

文件协同根据文件数据协同请求，将特定的文件数据交换到指定的其他信息协同节点，支持文本、二进制文件等不同文件数据的协同。

表4.2　数据库数据发送/接收接口参数表

参数名称	参数类型	可选/必选	参数含义
dbData	数据库协同数据对象DBExData	必选	需要通过数据库数据发送/接收接口发送/接收的数据集

（1）文件协同请求接口：将文件协同请求信息发送到管理中心，并在发送数据的信息协同节点与接收数据的信息协同节点之间建立会话。文件协同请求接口参数说明见表4.3（文件协同请求对象FileExRequest的XML Schema定义见附录C）。

表4.3　文件协同请求接口参数表

参数名称	参数类型	可选/必选	参数含义
Message	文件交换请求对象FileExRequest	必选	文件协同请求信息
deliveryMode	Int	可选	采用的传送模式，支持的传送模式至少应包括持续传送模式PERSISTENT和一次性传送模式NON_PERSISTENT
Priority	Int	可选	请求的优先级，共有10个优先级。0是最低优先级，9是最高优先级。默认的优先级是4
timeToLive	Int	可选	请求的存在时间，单位为毫秒，由发送方指定，0为无限制

（2）文件数据发送/接收接口：将需协同的文件数据发送给接收数据的信息协同节点。文件数据发送/接收接口参数说明见表4.4（文件协同数据



对象 FileExData 的 XML Schema 定义见附录 C)。

表 4.4　文件数据发送/接收接口参数表

参数名称	参数类型	可选/必选	参数含义
fData	文件协同数据对象 FileExData	必选	需要通过文件数据发送/接收接口发送/接收的数据集

4.3.3　物联信息传输

物联信息的传输从底层逐级向上分为传感网网关、支撑层和应用系统层三个层次。信息传输通信协议对应于 ISO/OSI 定义的 7 层协议中的应用层,在基于不同传输网络的传输层次之间提供交互通信。

物联信息的传输网络连接相对复杂,其基础传输层依据不同的传输网络分为基于 TCP/IP 协议的传输和基于非 TCP/IP 协议的传输两类实现方式。传感网网关和支撑平台之间采用基于 TCP/IP 协议的数据传输和基于非 TCP/IP 协议短消息数据通信两种方式;支撑平台与应用系统之间采用基于 TCP/IP 协议的数据传输方式。应用层依赖于所选用的传输网络,在选定的传输网络上进行应用层的数据通信,在基础传输层已经建立的基础上,整个应用层的协议和具体的传输网络无关。



1. 基于 TCP/IP 协议的传输

在存在有线或无线网络连接、能够进行基于 TCP/IP 协议的数据传输情况下,可以采用此类数据传输方式进行数据交换。此方式的使用建立在 TCP/IP 基础之上,常用的有通用无线分组业务(General Packet Radio Service,GPRS)、非对称数字用户环路(Asymmetrical Digital Subscriber Loop,ADSL)、码分多址(Code Division Multiple Access,CDMA)等。

1) 应答模式

完整的命令由请求方发起,响应方应答组成。具体步骤包括:请求方发送请求命令给响应方;响应方接到请求命令后应答,请求方收到应答后认为

连接建立;请求方发送请求的操作;响应方执行请求的操作;响应方通知请求方请求执行完毕,没有应答按超时处理;重复上述步骤,直到请求和响应的其中一方发送结束命令或超过一个适应性时长未收到对方数据;命令完成。

在基于 TCP/IP 协议的数据传输方式下,交互双方可以进行双向通信,网络层应兼容 TCP 和 UDP 两种传输协议。

2）超时重发机制

对应的超时主要包括以下四种情况：一个请求命令发出后在规定的时间内未收到回应,则认为属于请求回应超时。超时后重发,重发规定次数后仍未收到回应认为通信不可用,通信结束。超时时间和超时重发次数可以根据具体的通信方式和任务性质自定义。

请求方在收到请求回应(或一个分包)后规定时间内未收到返回数据或命令执行结果,则认为属于执行超时,命令执行失败,结束。

3）交互过程

信息的交互过程是指建立在应用层上的应用信息交互过程。在交互过程中,任意一方发送结束命令时,交互过程结束。

交互过程一般分为三个主要阶段：一是建立会话阶段,由发送方发起请求会话,接收方分配会话 ID 并确认会话;建立会话阶段涉及的报文格式有请求会话和确认会话;二是提交数据阶段,由发送方向接收方提交数据,接收方接收到数据后确认数据;提交数据阶段涉及的报文格式种类较多,包括确认数据所有格式;三是注销会话阶段,由发送方向接收方发送注销会话信息,接收方将此会话 ID 注销;注销会话阶段涉及的报文格式为注销会话。

4）数据结构

通信包采用二进制编码格式,未特别说明的均采用小端法机器存放数据方式,兼容基于 TCP 和 UDP 两种协议的数据传输。通信协议数据结构如图 4.6 所示,适用于 RS-232、RS-485 及 RS-422 等串口通信。

通信包是最小通信单元。在一次会话过程中,可以产生一次或多次传

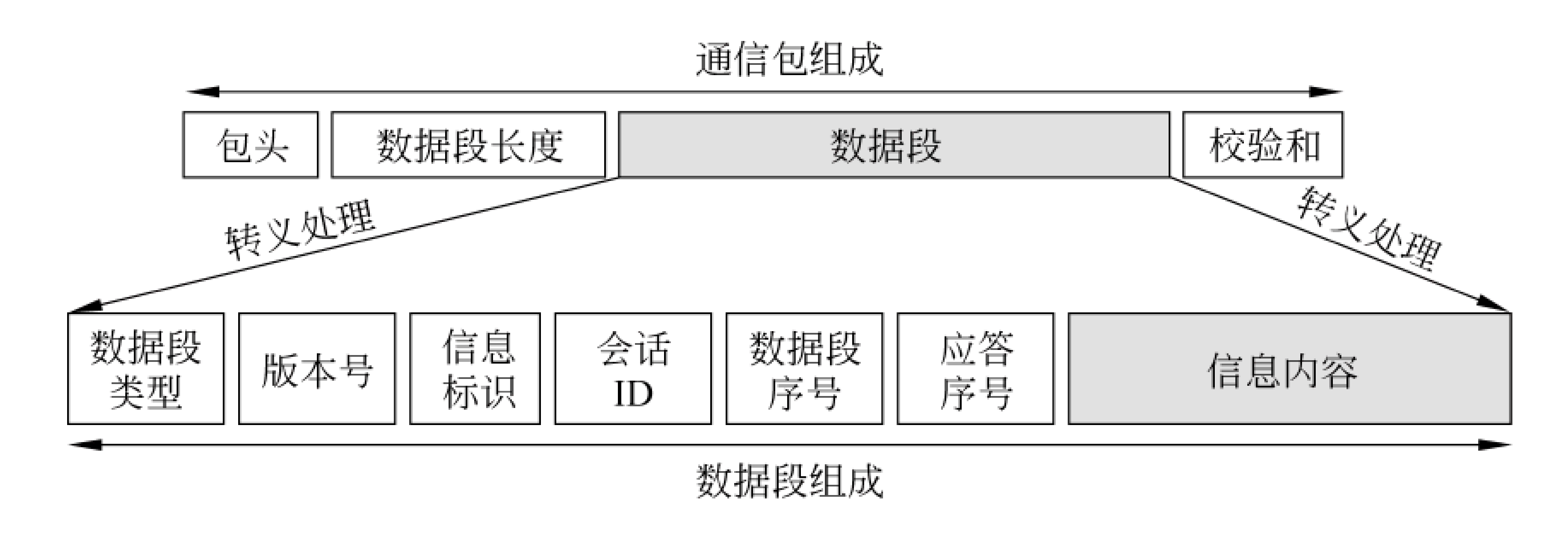


图 4.6 通信协议数据结构



输过程，即可以发送和接收多个通信包，但每次传输仅传输一个通信包。通信包结构组成如表 4.5 所示。

表 4.5 通信包结构组成

名 称	类 型	长 度	描 述
包头	字符	1 字节	固定为 0x7E，每一个通信包的开始字符
数据段长度	十进制整数	2 字节	数据段的数据长度
数据段	数据	0≤N≤4096	长度不固定，转义后的实际数据
校验和	十六进制整数	1 字节	数据段的校验结果，如果不为 0xFF 则表示出错

数据段部分为转义后的实际数据，在发送或接收数据时，为了避免数据中出现的字节与数据帧的标志性字符发生冲突，必须转义冲突数据值。数据 0x7E 和 0x7D 替换为两个字节：第一个字节是 0x7D，第二个字节是数据与 0x20 异或后的值。数据段经转义后得到的实际数据结构如表 4.6 所示。

表 4.6 数据段转义后的数据结构

名 称	类 型	长 度	描 述
数据段类型	十六进制整数	2 字节	固定为 0x0102，标示数据段所属类型，如物联数据
版本号	十六进制整数	1 字节	固定为 0x02，物联信息传输规范的生命期版本标示符，从 0x0001 版开始，每推进一个版本，数值增加 1，当前版本为 02 版

续表

名　称	类　型	长　度	描　　述
信息标识	十六进制整数	1 字节	用于标识信息的类别，包括参数命令、交互命令、数据命令和控制命令
会话 ID	十进制整数	4 字节	会话双方在一个会话过程中始终保持的唯一标示符
数据段序号	十进制整数	4 字节	以日时间作为通信包序号，日时间为从当天午夜 0 时起到当前时间的毫秒数，每毫秒不超过 1 个通信包
应答序号	十进制整数	4 字节	响应方所响应的数据段序号
信息内容	数据	0≤N≤4080	包括参数命令、交互命令、数据命令和控制命令的信息内容，本字段格式为可扩展格式，可通过其他补充规范对本字段格式进行补充定义

信息通信需要对涉及的交互命令信息内容进行数据结构的统一约束。交互命令是包括建立会话、断开会话及交互响应使用的一类信息，使用 0x00 到 0x3F 段的数据单元类型。

2. 基于非 TCP/IP 协议的传输

在不存在能够进行基于 TCP/IP 协议的信息传输，但能进行短消息或其他文本方式交互的情况下，可以采用基于非 TCP/IP 协议短消息数据通信的方式进行信息交换。此类方式的使用建立在相关通信链路上，常用的有公共电话交换网(Public Switched Telephone Network，PSTN)、短消息数据通信等。

1）通信模式及传输协议

在短消息数据通信方式下，按交互双方系统约定及短信网络架构，信息从传感网网关接入后，通过相应的短信或文本网关以短信形式进行单方向传输。

2）数据结构

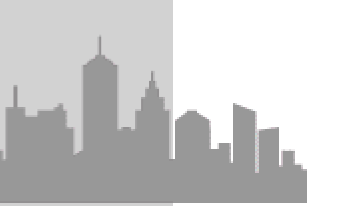
短消息数据通信方式的物联数据报文数据结构使用格式化的文本短



信。短消息数据通信格式的物联数据分为两种类型，参数命令和数据命令。每种类型的信息由多个数据项组成，数据项之间以“，”分隔。短消息数据结构如表 4.7 所示。

表 4.7　短消息数据结构

名称	类型	长度	描　　述
数据段类型	字符	6 字节	固定为“＃GZJR＃”
版本号	字符	可变长	固定为“2”，物联信息传输规范的生命期版本标示符，从“1”版开始，每推进一个版本，数值增加 1，当前版本为“2”版
信息标识	字符	可变长	用于标识信息的类别，包括参数命令和数据命令
信息内容	字符	可变长	包括参数命令和数据命令的信息内容，本字段格式为可扩展格式，可通过其他补充规范对本字段格式进行补充定义



3. 数据传输的循环冗余校验

循环冗余校验(Cyclic Redundancy Check，CRC)是一种数据传输错误检查方法。CRC 码两个字节，包含 16 位的二进制值。它由传输设备计算后加入到数据包中。接收设备重新计算收到消息的 CRC，并与接收到的 CRC 域中的值比较，如果两值不同，则有误。

具体算法如下：CRC 先调入一个值是全“1”的 16 位寄存器，然后调用一个过程将消息中连续的 8 位字节各当前寄存器中的值进行处理。仅每个字符中的 8 位数据对 CRC 有效，起始位和停止位以及奇偶校验位均无效。

CRC 校验字节的生成步骤如下：

(1) 安装一个 16 位寄存器，所有数位均为 1。

(2) 取被校验串的一个字节与 16 位寄存器的高位字节进行“异或”运算，运算结果放入这个 16 位寄存器。

(3) 将这个 16 位寄存器向右移一位。

(4) 若向右(标记位)移出的数位是 1，则生成多项式 1010 0000 0000 0001 和这个寄存器进行“异或”运算；若向右移出的数位是 0，则返回 3。

（5）重复（3）和（4），直至移出 8 位。

（6）取被校验串的下一个字节。

（7）重复（3）～（6），直至被校验串的所有字节均与 16 位寄存器进行“异或”运算，并移位 8 次。

这个 16 位寄存器的内容即 2 字节 CRC 错误校验码。校验码按照先高字节后低字节的顺序存放。

应用案例 2　城市实体的基础信息与编码管理

1. 基础信息属性

以某智慧城市的危险品监管为例，涉及运输、存储、销售等危险化学品流向的实时跟踪管理、综合保障、环境整治等多个环节，需要多领域、多部门、多层级、多主体的协同联动。

危险品监管领域的管理对象和感知设备的共性核心属性包括：名称、编码、所属业务、空间位置、负责方、频率、编/解码标准等。

管理对象和感知设备之间通过唯一的实体编码进行关联，如感知设备“GPS 全球定位系统”的基础属性中包含“管理对象编码”，与所感知的管理对象“货车”的唯一编码相对应。

根据不同的业务特定，物联实体还具备一系列扩展属性。

（1）管理对象的基础信息：如“仓库”具有“库房间数、仓库面积、灭火器数、消防水源、限制存箱量、限制存药量”等扩展属性，“批发单位”具有“安全管理人员数量、仓库保管和守护人员数量、运输车辆数量” 等扩展属性，“消防车”具有“车辆牌号”的扩展属性，“自动气象站”具有“建设时间”等扩展属性。

（2）感知设备的基础信息：如“温度计”具有“所属仓库、所属库房、所属房间、型号、最小刻度、最大刻度” 等扩展属性，“空气质量传感器”具有“量程、工作温度、存储温度、加热阻抗、加热电压”等扩展属性。



2. 物联实体编码与管理

管理对象和感知设备编码整体结构采用前段码与后段码组合的形式，前段码与后段码用“.”分隔；传感网网关编码统一进行自动赋码。前段码为定长码，由版本码 2 位(编码管理部门发布)、类别码 1 位(管理对象、感知设备、传感网网关)、单位编码 X 位(如组织机构代码)组成，由编码管理部门统一管理；后段码为非定长码，由业务部门依据业务规则或现行标准自行编制和管理。

智慧城市的各相关部门将管理对象、感知设备和传感网网关的编码及其他基础信息在统一的平台上进行注册登记。如管理对象、感知设备和传感网网关基础属性发生变更时，各部门在统一的平台上进行变更登记。注册后的管理对象、感知设备和传感网网关基础信息通过统一的平台为权限内的用户提供查询服务。



3. 管理对象编码(采用十进制)

以某智慧城市的交通管理车为例，管理对象的编码结构如下：

(1) 前段码结构：包括版本码、类别码和单位编码三个部分。根据前段码管理部门发布的版本信息，当前管理对象编码版本为“01”；管理对象的类别码为“1”；车辆的管理单位为交通部门，组织机构代码为“110018”。因此，交通管理车的前段码为“011110018”。

(2) 后段码结构：按照有关规定，机动车分类编码应按照 GB 918.1—1989 规定编制，但该标准仅规定了类型代码，不能保证单个机动车代码的唯一性，顺序码不可省略，故后段码编码结构为“现有标准分类编码＋顺序码＋校验码”的形式。按照 GB 918.1—1989 机动车分类编码为 3 位 10 进制数字编码，交通管理车大类为机动车，大类编码为 1；中类为特种车，中类编码为 5；小类为交通管理车，小类编码为 5，因此交通管理车的标准分类编码为“155”。假设该市交通管理车不超过 10 万辆，编码主管部门发给交通管理车 1～10 万的顺序码，即交通管理车占用的顺序码应不少于 5 个码位。

结合前段码和后段码，该市交通管理车的编码为：011110018.155000001～011110018.155100000。

4.感知设备编码（采用十进制）

以某智慧城市公安部门的交通视频探头为例，感知设备的编码结构如下：

（1）前段码结构：包括版本码、类别码和单位编码三个部分。根据前段码管理部门发布的版本信息，当前感知设备编码版本号为“01”；感知设备的类型码为“2”；视频探头的管理单位为公安交通管理部门，组织结构代码为“110067”。因此，交通视频探头的前段码为“021110067”。

（2）后段码结构：因交通视频探头的出厂标准编码已经为唯一值，故省略顺序码，后段码编码结构为“感知设备出厂标准编码＋校验码”的形式。交通视频探头的出厂标准编码共14位，形式为“XXXXXXXXXXXXXX”。

结合前段码和后段码，交通视频探头的编码为：012110067.XXXXXXXXXXXXXX。

目前，大部分已颁布执行的代码和编码标准针对专业、行业应用范畴，因此嵌入后段码时，可以根据管理需要参照选用标准的编码规则，仅截取部分有效的码段和编码。



应用案例3　智慧城市危险品监管的实时信息接入

以某智慧城市的危险品监管为例，物联实时信息涉及安全监管、环境保护、气象、消防、卫生、市政管理、城管执法、民防、交通、公安等多个领域和部门，各信息源的实时信息、流转模式、获取频率及对应的感知设备如表4.8所示。

根据实际管理需求，视频信息主要采用直接接入和系统对接两种模式，车辆位置信息主要采用接入转发模式，实时监控信息主要采用前置交换模

式，预警预报信息主要采用 Portlet 嵌入模式。所有物联信息均按照统一的接口标准和技术规范在信息源、信息提供方、基础支撑平台（中心管理端）、信息需求方之间有序流转。通过信息流转标准的约束，可以实现城市管理不同领域的信息协同，有效支撑领域精细化管理和科学决策。

表 4.8　危险品监管领域的物联实时信息接入

信息来源	实时信息	流转模式	获取频率	感知设备
安全监管部门	危险品运输车出入库信息	前置交换	5 分钟	RFID
	仓库报警信息（温度、湿度、越界等）	前置交换	5 分钟	温度传感器、湿度传感器、红外探测器等
	仓库实时状况信息	直接接入	实时	标清摄像头
	销售实时状况信息	直接接入	实时	标清摄像头
环保部门	噪声信息	前置交换	5 秒钟	噪声传感器
	大气成分含量信息	前置交换	1 小时	可吸入颗粒物传感器
气象部门	实时气象信息（风速、风向、气压、雨量等）	前置交换	2 分钟	风速传感器、风向传感器、气压传感器、降水传感器等
	气象预报信息	Portlet 嵌入	1 小时	—
消防部门	消防车位置及运动轨迹信息	接入转发	5 秒钟	北斗卫星导航系统
卫生部门	救护车（120、999）位置及运动轨迹信息	接入转发	5 秒钟	GPS 全球定位系统
城管执法部门	城管执法车位置及运动轨迹信息	接入转发	5 秒钟	GPS 全球定位系统
民防部门	视频监控信息	系统对接	实时	高清摄像头
交通部门	危险品运输车位置及运动轨迹信息	接入转发	5 秒钟	GPS 全球定位系统
公安部门	视频监控信息	系统对接	实时	高清摄像头

第5章 智慧城市多源信息协同的自适应模式

信息协同模式是多源信息协同体系的核心部分，与信息协同评价之间形成系统内部信息流转的闭环流程，根据信息协同网络的测度结果进行信息协同模式的自适应优化。本章重点介绍多源信息协同的模式及其自适应的智能化过程。根据智慧城市的大数据特征和应用需求，分析不同业务需求下的信息协同模式，通过改变信息的流向对传统模式下的信息流转进行优化；根据信息协同不同阶段的特点构建信息流转的自适应进程，实现信息协同技术流程与业务流程的分离，最终实现信息在城市系统中智慧地流转。

5.1 智慧城市系统下的信息特征与流向分析

智慧城市系统中流转的信息可以归纳为业务信息和日志信息两大类。其中业务信息主要包括普通的业务信息、来自于物联终端的实时感知信息、视频信息等三种类型，日志信息主要指信息流转过程中产生的事件信息和状态信息。普通业务信息一般可以按业务部门的领域职能和业务流程进行划分；视频信息在信息载体和特征上具有自身的特殊性；对实时感知信息的分类是关键，这就需要对实时感知信息的特征进行合理的描述。

信息分类是信息特征分析的基础。多维信息的分类需要在基础指标基础上构建多维度指标体系的关联关系，在分类过程中遵循基础分类与辅助分类相结合、求大同存小异、用语规范性与灵活性相结合、具有层次性和可扩展性等基本原则。多维信息的分类方法如图5.1所示。

智慧城市的物联信息特征可以按照人、地、事、物、组织、领域、时间七个维度进行描述，如表5.1所示。

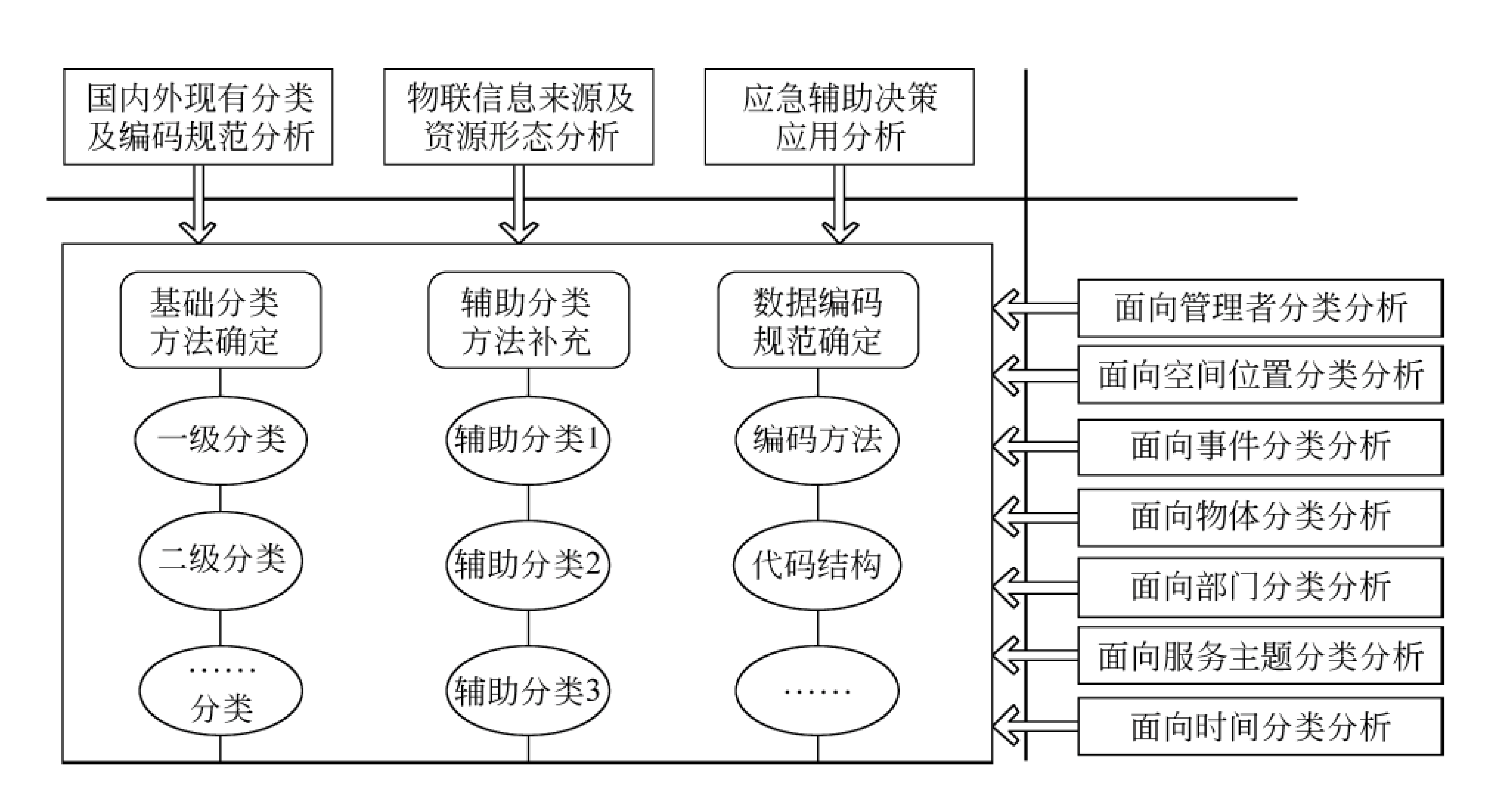


图 5.1　多维信息分类方法

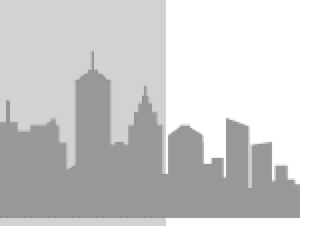


表 5.1　智慧城市的物联信息特征

信息维度	特 征 描 述	核 心 因 素
人	信息的组织、存储与应用	信息的管理者
地/空间	反映城市体征状态的关键空间实体“天地一体化”	空间实体位置
事件	信息与事件的关联关系	应急事件
物	信息的对象主体	感知设备和管理对象
组织	横向跨部门、纵向跨层级	信息的所属部门
领域	信息的业务领域	服务主题
时间	信息的过程状态(动态)和结果状态(静态)	信息产生/需求的频率

(1) 按“人”的信息维度描述：如“提供方的信息”、“使用方的信息”、“管理方的信息”等。

(2) 按“地/空间”的信息维度描述：如“地下管线信息”(地下)、“公交线路信息”(地表)、“空气质量信息”(地上)等。

(3) 按“事件”的信息维度描述：如“安全防汛信息”、“森林防火信息”、“扫雪铲冰信息”等。

(4) 按"物"的信息维度描述：如"烟花爆竹仓库的温度信息"、"应急抢险车辆的位置信息"、"电梯的运行状态信息"等。

(5) 按"组织"的信息维度描述：如"城管部门信息"、"市政部门信息"、"水务部门信息"等。

(6) 按"领域"的信息维度描述：如"气象领域专题信息"、"交通领域专题信息"、"城市生命线领域专题信息"等。

(7) 按"时间"的信息维度描述：如"自动气象站的位置信息"(静态)、"气象云图的实时信息"(秒级动态)等。

智慧城市的大数据特征突出体现对海量实时感知信息的共享与整合。如何让城市系统运行的状态信息、事件信息和决策信息以最快的速度、最优的模式(路径)从信息源流向需求方，决定了城市管理决策的效率，进而直接影响城市安全运行、公共服务和应急决策的水平。智慧城市运行管理的信息流向分层架构如图 5.2 所示。

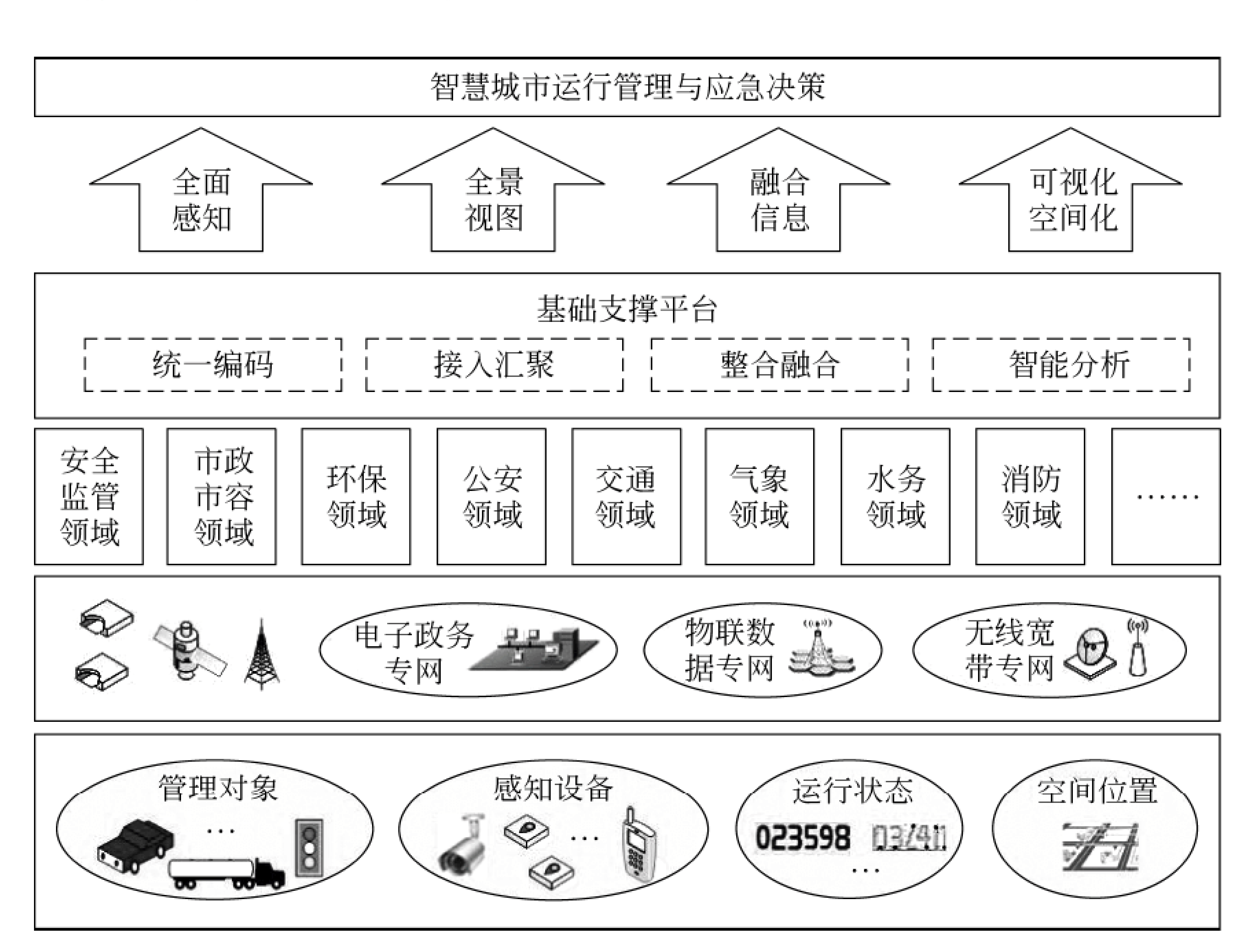


图 5.2　智慧城市运行管理的信息流向分层架构

通过城市运行管理部门应用的信息交换和领域应用的信息接入，将多源异构的数据与人口、法人、地理空间等基础信息资源进行整合，最终形成的融合信息进一步应用于城市运行管理的各个部门和领域业务，实现信息流转的闭环管理(见图 5.3)。

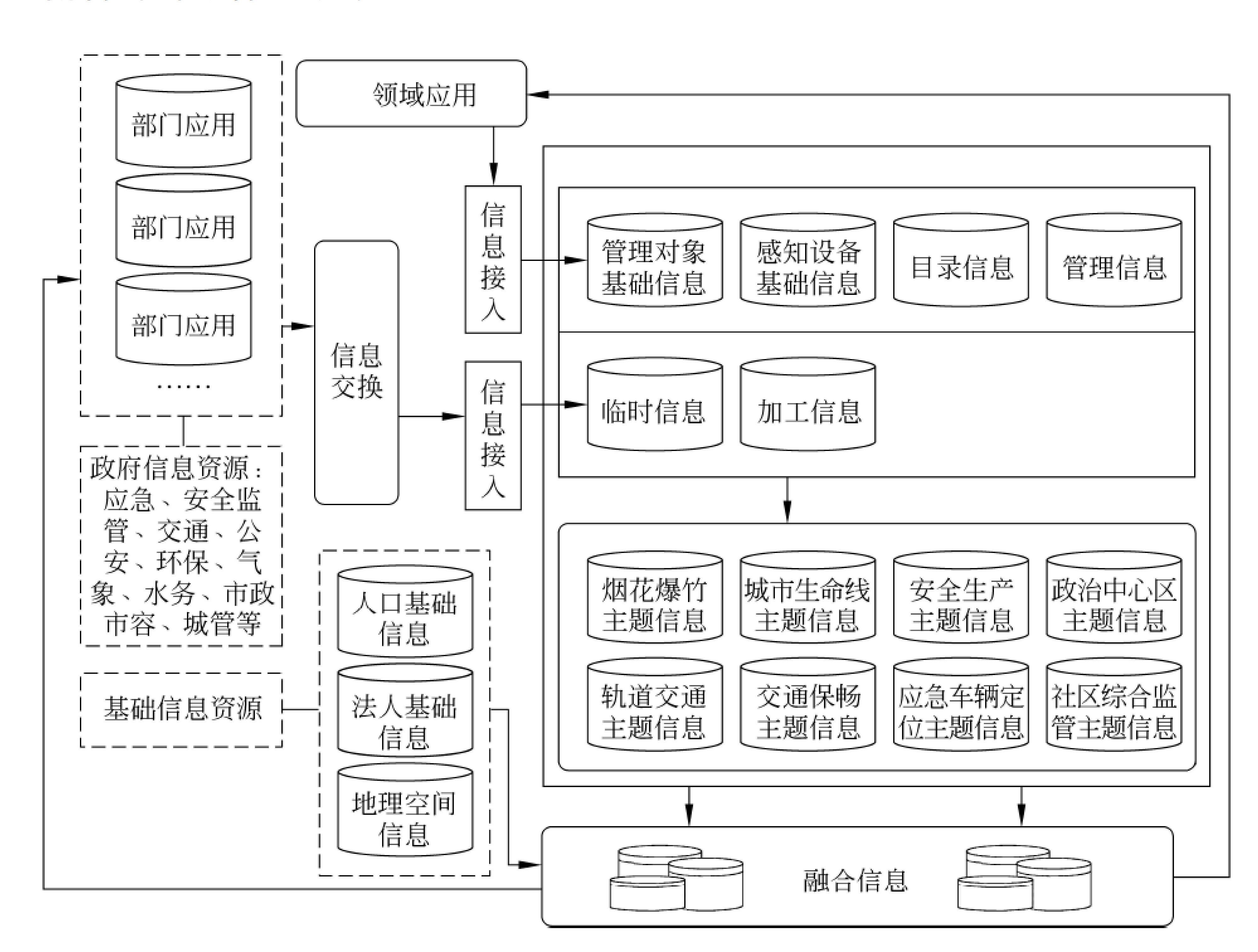


图 5.3　智慧城市运行管理的信息流转闭环模型

应用案例 4　城市基础运行领域的物联实体分类

某智慧城市基础运行管理领域的管理对象分类如表 5.2 所示。

感知设备分类如表 5.3 所示。

表 5.2　某智慧城市基础运行管理领域的管理对象分类

一级分类	二级分类
车辆	消防战斗车、重型专项作业车、中型专项作业车、消防指挥车、泡沫消防车、气防车、备用车、指挥车、后勤车、消防战斗车、干粉泡沫联用车、应急指挥车、抢险车、发电及发电辅助用车、吊车、拖车、挖掘机、工程抢险车、高压注浆机、钻机、应急指挥通信车、铲雪车、防洪车、货车、专用客车、拖板车、振动压路机、轮式挖掘破碎机、拖车式电站、防洪车、铲雪车、洒水车、应急供水车、城市管理综合执法保障车辆、环境监测车、公交车辆、出租车辆、长途巴士、通信车辆、救护车、垃圾车、渣土运输车、清扫车……
道路桥梁	桥梁、重点桥区、市内道路、高速公路、重点路面、高边坡……
交通站点	公交车站、汽车总站、长途客运站、轨道交通车站出入口、轨道交通换乘通道、轨道交通站台……
供电设施	电力联络线、输电线路、固定施工场所、非民用用电设施……
供水设施	水源地、水厂、供水管网、水箱、二次供水水箱、供水水表井下、车间或源库……
排水设施	污水处理厂、再生水厂、污水排水管网、雨污水泵站、再生水输配管网、再生水提升泵站、废水总排口、排入水口、污水处理站、废水总排口……
供热设施	热热源厂、锅炉、热力管线、小室、热力站……
燃气设施	燃气门站、燃气调压站、非民用燃气设施……
通信设施	光缆、有线电视……
人防工程	应急避难场所、民防工程、消防中控室、民防高点、疏散通道、安全出口……
监测站点	环境监测点、噪声监测点、气象监测站、高速公路气象监测站、高山气象站……
危险化学品	气瓶、气瓶间、危险地段、烟花爆竹批发单位仓库、危化品库房、烟花爆竹销售点、加油站、加油区、卸油区、地下灌区、天然气调压站、液氧站、危化品工作区、原料灌区、燃气锅炉房……
建筑工程	公园、主景区广场、休闲娱乐场地、景区制高点、建筑物高点、学校……
特种设备	电梯、起重机械、大型游乐设施……
厂矿设施	露天采场、炼油厂、辅助车间、排土场、井下、尾矿库、矿山地面、矿山外围、水泵、风门、皮带、道岔、锅炉、厂房车间、设备控制室、总配电室……



表 5.3 某智慧城市基础运行领域的感知设备分类

一级分类	二级分类
传感器	温度计/温度传感器、湿度传感器、红外探测器、空气质量传感器、噪声传感器、风速传感器、风向传感器、气压传感器、降水传感器、力/重量传感器、加速度传感器、流量传感器、位移传感器、尺度传感器、浊度传感器、密度传感器、黏度传感器、硬度传感器、热流传感器、热导率传感器、图像传感器、色传感器、激光传感器、磁场强度传感器、磁通密度传感器、电流传感器、电压传感器、电场传感器、超声波传感器、声压传感器、声表面波传感器、射线传感器、辐射剂量传感器、气体分压传感器、PH 传感器、成分传感器、离子活度传感器、水分传感器、心电传感器、血氧传感器、体温传感器、血压传感器、微生物传感器、细胞传感器、组织传感器、免疫传感器、酶传感器、生命探测装置……
视频监控设备	模拟摄像头、标清摄像头、高清摄像头……
定位设备	CNSS(区域性有源三维卫星定位与通信系统、北斗卫星导航系统)、GPS(全球定位系统)、基站定位……
射频识别设备	射频识别只读器、射频识别读写器……

5.2 城市系统下的多源信息协同模式

5.2.1 多源信息协同的应用模式

传统城市系统下的多源信息协同以信息组织间的信息共享为主，信息协同模式比较单一，主要是中心管理模式。智慧城市系统下的多源信息协同根据其大数据的特征及科学决策对信息实时性的要求，需要综合采用多种新的模式，主要分为信息共享(信息组织间的信息协同)和接入转发(实时感知信息的协同)两类。

信息共享模式主要包括中心管理、中心转发、节点对接和领域应用四种模式(见表 5.4)。

表 5.4　智慧城市运行管理的信息共享模式

模式类型	标　识	模式描述	协同关系
中心管理模式	PCP	中心管理、不落地	一对一
中心转发模式	PCN/NCP	中心管理、落地	一对多、多对一
节点对接模式	P2P	自行管理、不落地	一对一
领域应用模式	P2N	二级中心管理	一对多

1. 中心管理模式(PCP)

PCP 模式由管理中心来配置端到端的流程，包括控制流、消息流和数据流，对端节点业务应用提供服务。参与协同的节点通过前置机方式接入管理中心，流程的配置和控制(包括信息在流转过程中的传输和处理)由管理中心完成。

PCP 模式应用于信息组织之间一对一共享的情况，协同信息原则上不在交换中心落地，适合业务需求明确(一定时期内无频繁变化)、信息协同能力相对薄弱的信息组织。PCP 模式的信息流向如图 5.4(a)所示。

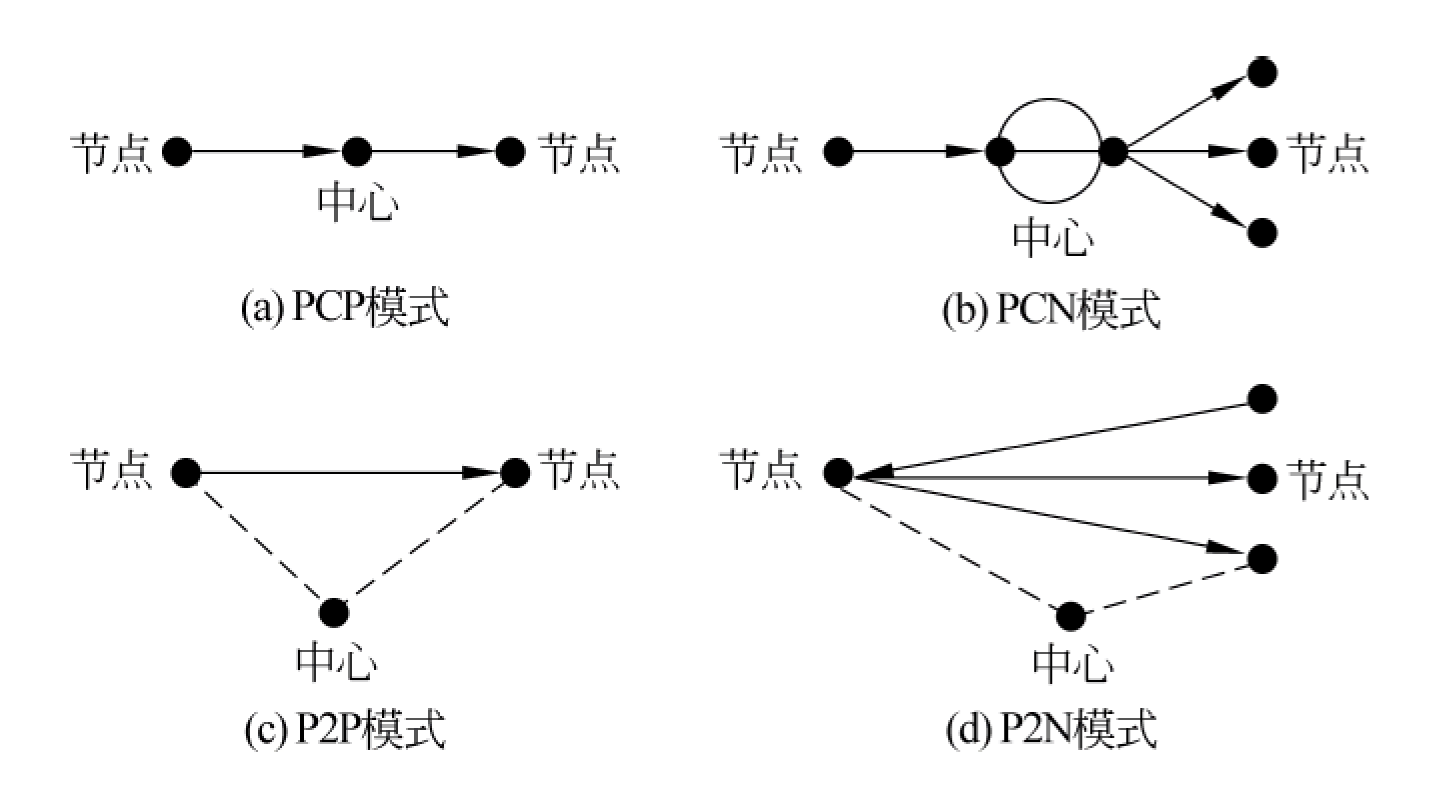


图 5.4　信息共享模式的信息流向示意图

2. 中心转发模式(PCN/NCP)

PCN/NCP 模式将信息协同流程分成两个阶段，从发送方节点到管理中心，再由管理中心到需求方节点。中心端根据业务需求进行必要的数据整

合与转化。该模式包括三种情况：

（1）PCN 模式：由发送方节点配置信息发送流程，中心端配置信息转发流程和信息接收流程。

（2）NCP 模式：由中心端配置信息发送流程和信息转发流程，需求方节点配置信息接收流程。

（3）PCN 与 NCP 结合的模式：由双方节点分别配置信息发送流程和信息接收流程，中心端配置转发流程进行对接，完成信息协同的全过程。中心端制定标准，配置节点间流程（不含节点内部流程），中心将流程（服务接口）授权（可视）给节点部门用于对接工作，各参与业务的部门配置本部门节点流程（服务）与中心对接。

PCN/NCP 模式应用于信息组织之间一对多共享的情况，协同信息需要在中心端落地或加工（整合）后转发。该模式中信息提供方与信息需求方不明确，需要数据进行不确定多向流转或不确定并发流转。PCN 模式的信息流向如图 5.4(b)所示。



3. 节点对接模式（P2P）

P2P 模式是在两个节点间进行直接信息协同的模式，管理中心不参与流程与业务的管理和控制。在管理中心的授权下，提供方节点和需求方节点自行配置信息发送和接收的流程，管理中心对信息协同过程进行监管。

P2P 模式应用于业务需求频繁、信息协同双方技术体系成熟（信息协同程度较高）的信息组织之间。P2P 模式的信息流向如图 5.4(c)所示。

4. 领域应用模式（P2N）

P2N 模式以应用领域的业务牵头部门（主业务节点）为二级管理中心，组织领域范围内的相关部门开展信息协同。由主业务节点制定标准，配置业务节点流程（服务），并将流程（服务接口）授权（可视）给相关参与部门；各参与部门配置本部门节点流程（服务）与主业务节点流程（服务）对接。

P2N 模式应用于业务重要性和独立性强、领域相关部门间有明确的业

务主从之分、牵头部门技术体系成熟(信息协同程度较高)的情况。该种模式改变了传统的“1个管理中心＋N个应用节点”的结构,形成了“1个管理中心＋M个领域应用中心＋N个应用节点”的二级(或多级)信息协同体系,将应用和管理的功能从中心端适当下放,利于信息在跨层级间的灵活架构和快速流转。P2N模式的信息流向如图5.4(d)所示。

接入转发模式主要通过物联数据专网和接入移动网络,将感知设备的实时感知信息统一接入到交换中心,由中心对感知信息网络包进行解码处理,获取接入网关的ID等信息,进行白名单验证和部门转发关系映射。接入转发的信息流转主要包括直接转发、存储转发和存储分发三种模式(见表5.5)。

表5.5　接入转发模式

模式类型	模式描述
直接转发模式	将验证通过的网络包直接转发至信息需求方
存储转发模式	直接转发的同时,同步存储入库,为多源信息的整合融合提供支撑
存储分发模式	直接转发的同时,根据需求同步分发到其他相关的信息需求方

接入转发的信息流向如图5.5所示。

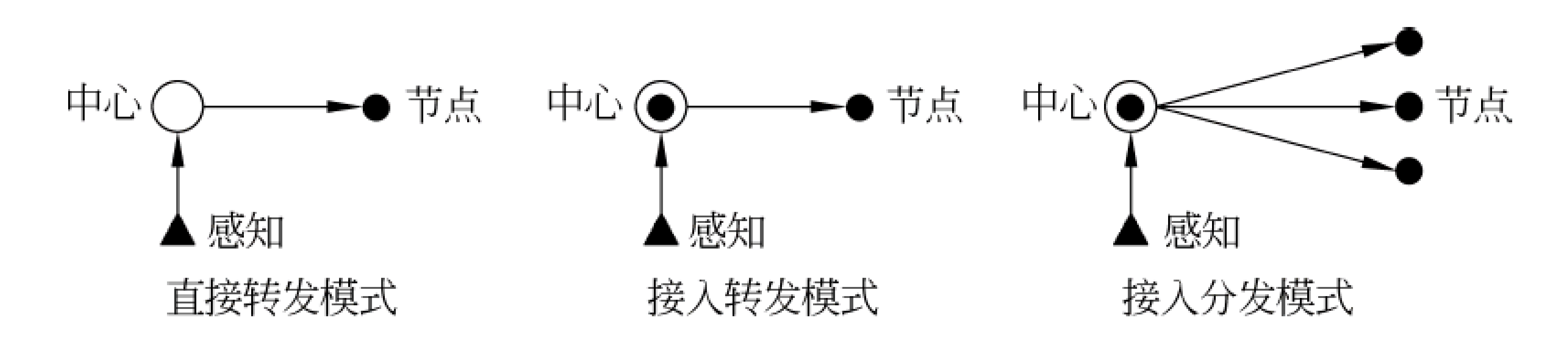


图5.5　接入转发模式的信息流向示意图

接入转发模式改变了传统的城市系统下的信息流向,信息不再全部由源头部门获取后通过交换中心共享,而是根据决策需求选择由中心直接从信息源获取后转发,一方面利于提高实时信息的流转效率,另一方面利于对跨领域信息的关联整合,提高决策支持的水平。

5.2.2 多源信息协同模式的流程分析

信息协同模式通过信息协同流程实现，主要包括多源信息协同的数据交换流程、管理监控流程、运行管理流程、事件通知流程、信息同步流程、节点部署流程和业务建模运行流程七种类型，下面对各类流程分别进行分析。

1. 多源信息协同的数据交换流程

（1）数据从信息提供方的业务系统或业务数据库中进行提取，通过桥接系统置于前置数据库中（或直接调用交换中间件提供的 API 接口），从而进入交换流程，如图 5.6 所示。

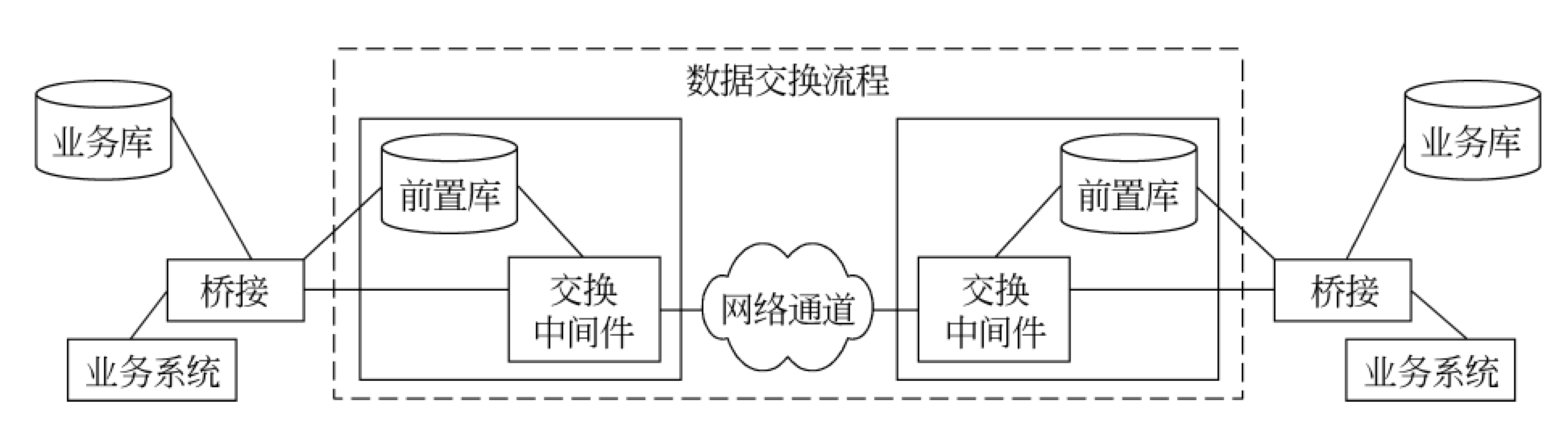


图 5.6　数据交换流程



（2）交换中间件从前置数据库中监控到数据的变化（或调用 API）后，待交换的数据被处理和发送。

（3）通过数据交换流程到达信息需求方的交换中间件，经处理后进入需求方前置数据库，再经桥接系统进入需求方的业务系统（或直接经桥接系统进入需求方业务系统）。

2. 多源信息协同的管理监控流程

1）管理中心的管理监控

管理中心的事件通知流程和管理监控流程如图 5.7 所示。

管理中心的管理监控主要包括三个方面：一是信息协同中间件的运行管理（如中间件的远程启动、停止等）；二是数据处理流程（业务）调度配置；三是数据传输连接器配置（对连接器 JDBC、JMS、FTP、HTTP（s）、

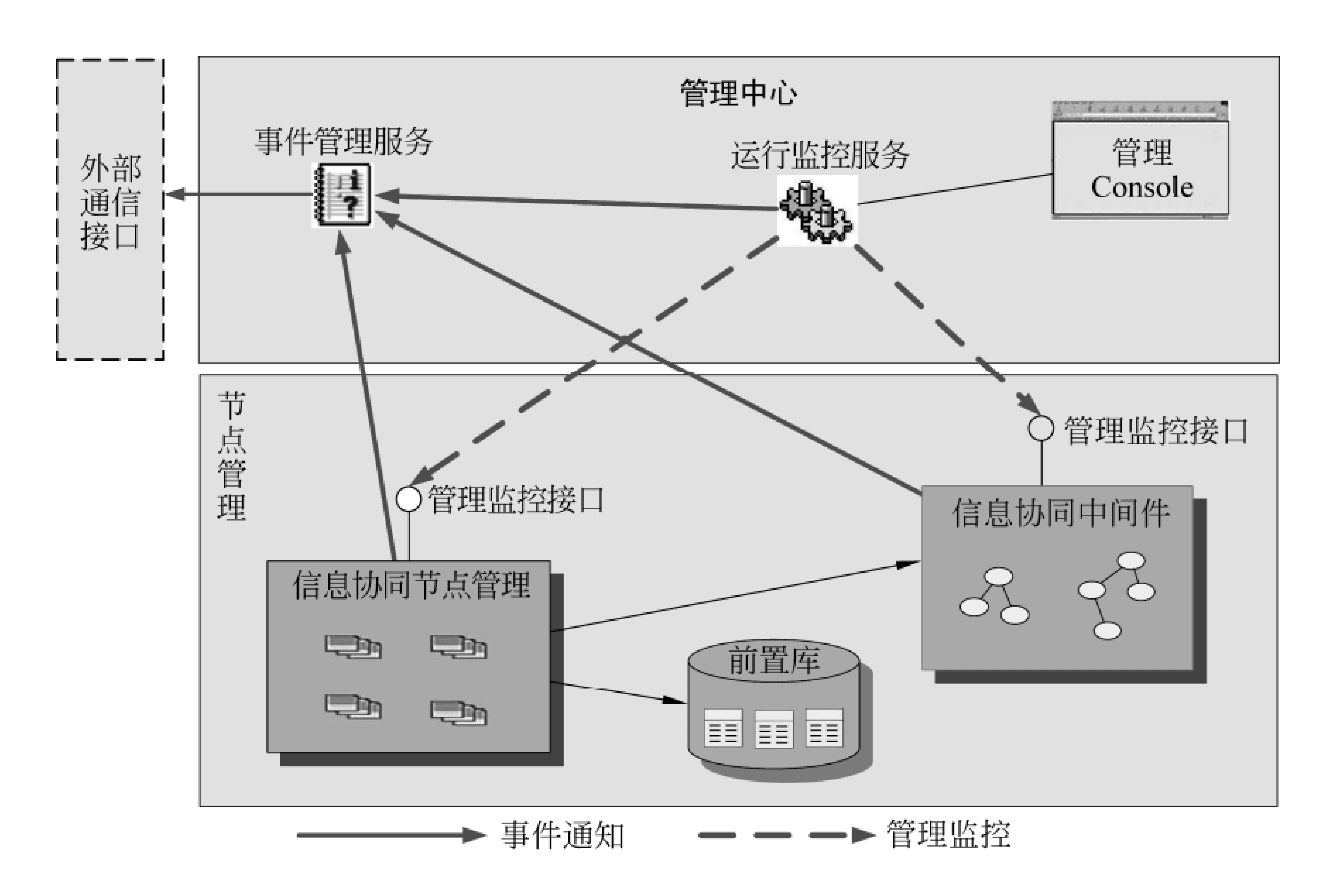


图 5.7 管理中心的事件通知流程和管理监控流程

WebService、REST 等的参数配置管理)。

其中,数据处理流程主要包括三种调度策略:一是实时策略,基于事件驱动、动态产生和控制的信息协同方式;二是定时策略,由定时器根据预设的定时策略产生和控制的信息协同方式;三是手工策略,由用户手动产生和控制的信息协同方式。调度策略与业务需求及建模存在密切联系。

2) 信息协同节点的管理监控

管理中心通过节点管理器实现对信息协同节点、信息协同中间件及前置数据库的运行状态监控。管理监控对象如图 5.8 所示。

信息协同节点的运行状态监控主要包括四个方面:一是节点服务器运行监控,如硬盘占用、内存占用、CPU 等情况,设置报警阈值,超过阈值后实时报警;二是信息协同中间件运行监控,采用轮询方式,监控中间件的运行状态,如就绪、运行、停止、挂起、异常等;三是数据处理(业务)流程监控,采用实时交互方式,通过中间件的 JMX 接口监控部署数据处理流程(及组件)的运行状态,如就绪、进行、处理完成、回执确认、处理中断、异常

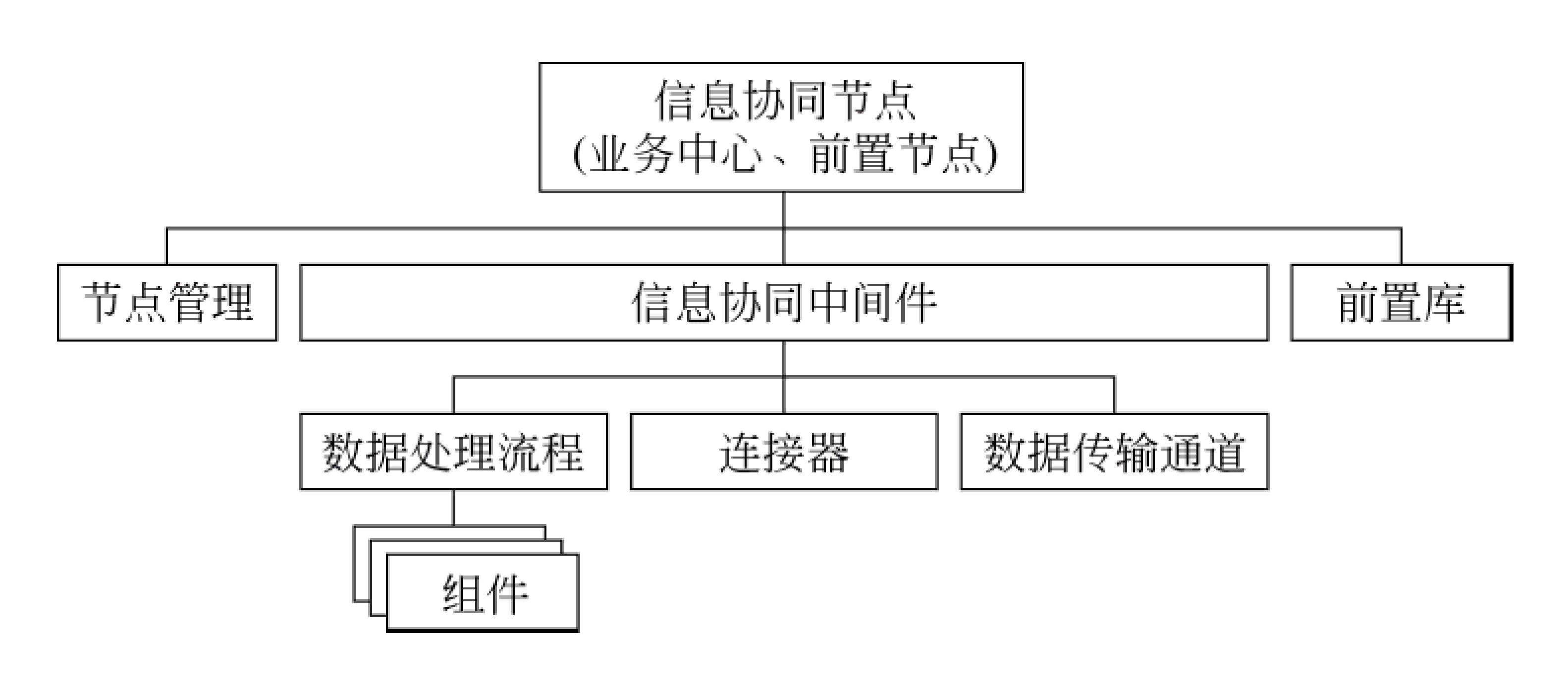


图 5.8 管理监控对象

等；四是节点数据传输通道监控，采用实时交互方式，通过中间件的 JMX 接口监控数据传输通道的有效性，通过 JMS 方式监控队列中待处理消息的情况。

信息协同中间件主要监控数据处理流程、连接器、数据传输通道及相关组件。各监控对象的属性包括：名称、类别（数据处理流程、连接器、数据传输通道、组件等）、所属节点、所属业务（包含若干任务）、所属任务（对应某项业务）等。



3. 多源信息协同的运行管理流程

信息协同过程中进行强日志管理，并提供基于 Web 的日志查询。对不同的信息协同流程日志进行汇总，实现按时间范围、流入/流出、提供/需求、信息载体等多维度的统计。日志查询流程和汇总流程如图 5.9 所示。

4. 多源信息协同的事件通信流程

信息协同节点（节点管理器、中间件、数据处理流程、组件等）在运行过程中与管理中心之间进行双向的事件通信，管理中心对事件消息进行统一管理和监控。事件通信流程如图 5.10 所示。

5. 多源信息协同的信息同步流程

实现管理中心到业务中心的基础信息同步，从而实现业务中心的独立运行。

图 5.9 日志查询流程和汇总流程



6. 多源信息协同的节点部署流程

在信息协同节点部署节点管理器和信息协同中间件，信息协同中间件在节点管理器上进行注册并提供相应的功能脚本，注册信息写入节点管理器的配置文件进行管理。

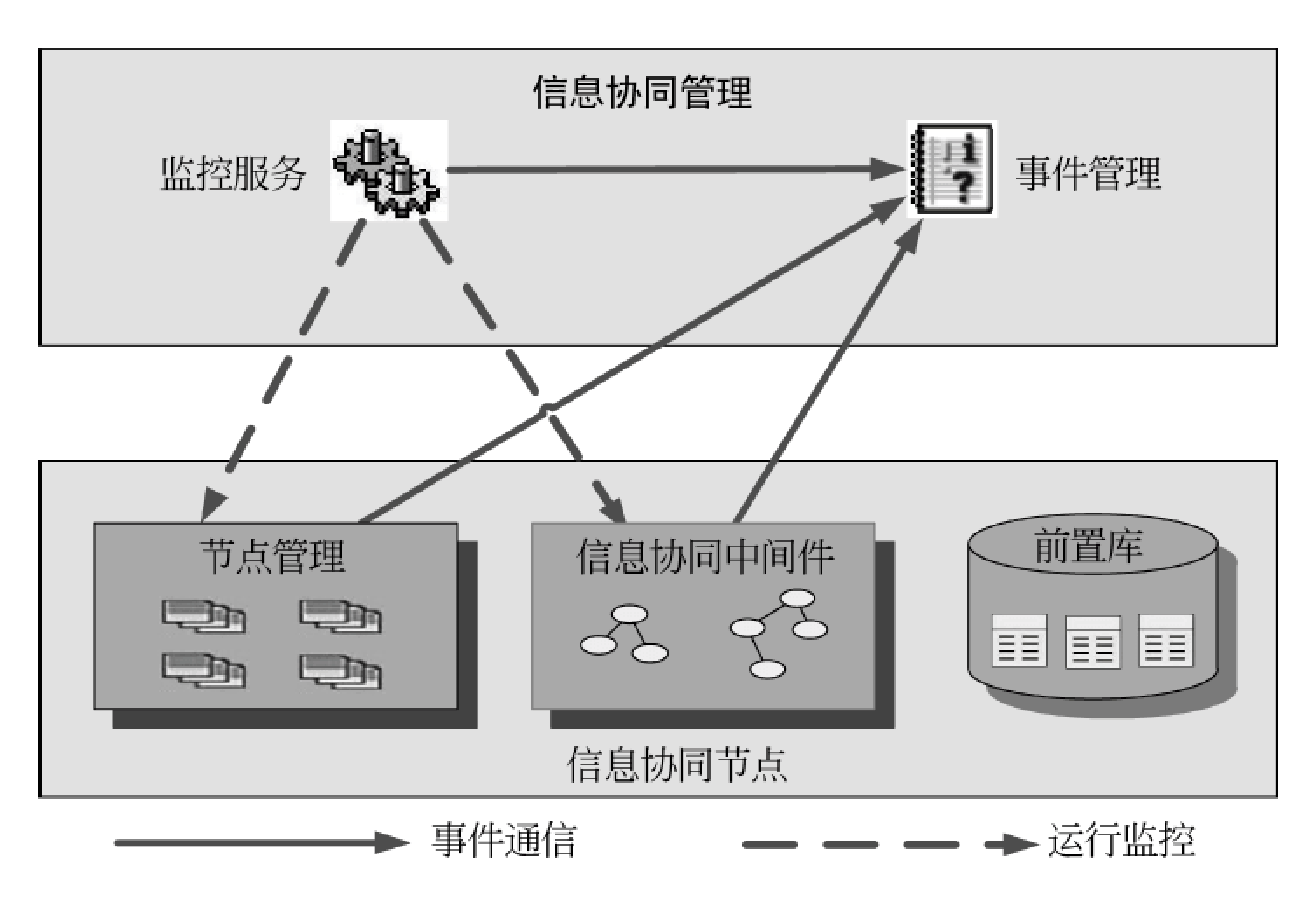


图 5.10 事件通信流程

7. 多源信息协同的业务建模运行流程

信息协同的业务建模运行流程主要包括四个阶段：一是业务规划及业务建模，在管理中心注册业务，规定业务域名称（在建模及 JMX 接口实现中使用）；二是数据交换建模，按照业务流程和信息协同需求进行建模，定义数据的抽取、转换、校验、加载等，并定义异常处理流程，保证处理异常能够及时被捕获并通知管理中心；三是部署，将所建好的数据处理流程发布到远程前置节点并进行测试，将该流程的相关管理监控接口注册到管理中心并进行测试，数据处理流程的部署和卸载均在管理中心事件管理器中进行事件通知；四是运行管理监控，通过管理中心经由所注册的接口进行管理监控。业务建模运行流程如图 5.11 所示。

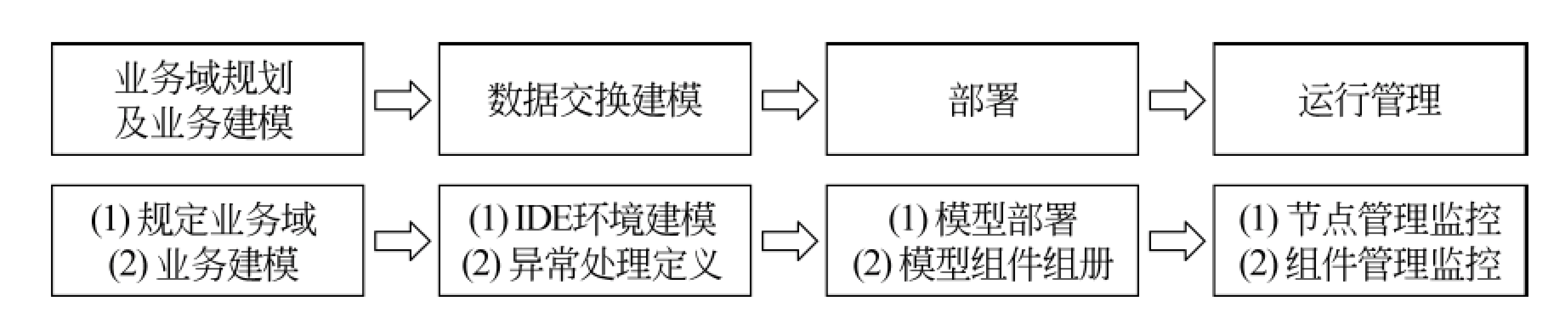


图 5.11 业务建模运行流程

5.3 城市系统下的多源信息协同自适应过程

5.3.1 多源信息协同的阶段

根据城市系统下的信息流转方式和特点，信息协同的全流程分为信息启动、信息流转、信息到达三个阶段，各阶段的信息整合方式和整合内容可概括为表 5.6 的形式。

表 5.6 城市系统下的信息协同阶段

信息协同阶段	信息整合方式	信息整合内容
信息启动阶段	聚类整合	信息的分类
信息流转阶段	加工整合、关联整合	信息的转换和关联
信息到达阶段	业务整合	信息技术流程与业务流程的对接

1. 信息启动阶段

根据信息的分类标准，对信息进行多维度的精细化分类。不同维度的分类信息之间存在交叉，分类的维度和粒度关系到流转阶段和到达阶段的信息定位与整合。在信息分类的基础上，根据提供方的业务需求和业务规则选择不同的共享方式，主要包括公开、普适共享、有限共享、特定共享和专用五种情况，如表 5.7 所示。

表 5.7 城市系统下的信息共享方式

共享边界	面 向 对 象	共享关系	信 息 实 证
公开	全社会	—	气象部门的天气预报信息
普适共享	所有政府部门	一对多	地理空间的基础图层信息
有限共享	特定领域或特定事件相关的部门	一对多	环保部门的环境噪声信息
特定共享	具体业务需求指定的特定部门	一对一	公安部门的高清视频信息
专用	部门内部使用	—	依法保密的相关涉密信息

2. 信息流转阶段

在流转阶段对信息的内容、数量、阈值等的规范性和合理性进行监控，并根据需求进行信息的加工转换和多源信息间的关联整合。信息流转阶段是信息协同过程中的核心环节，主要包括信息交换、信息接入和信息整合三个方面。

信息交换模型如图 5.12 所示。

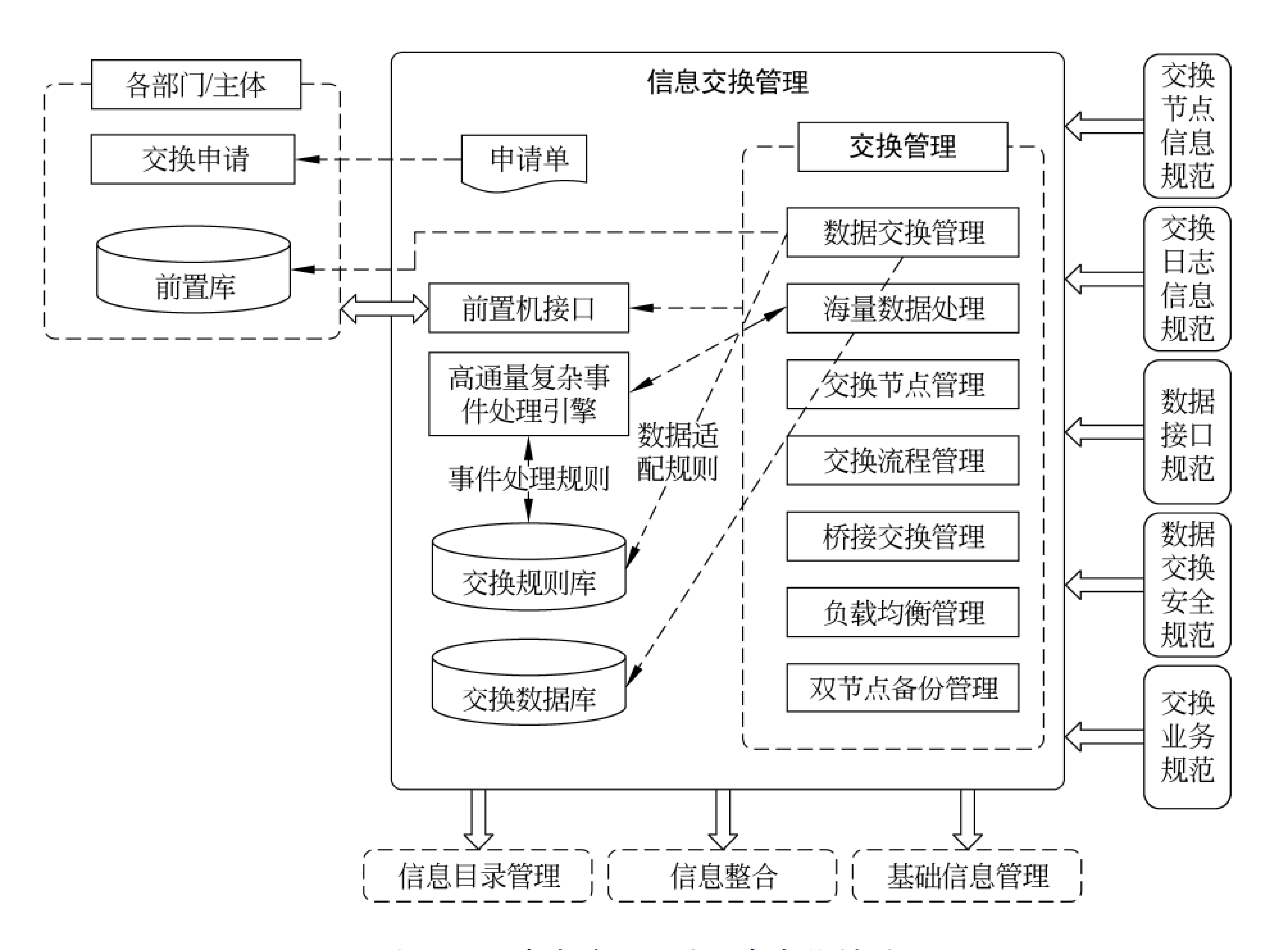


图 5.12　城市系统下的信息交换模型

信息接入模型如图 5.13 所示。

信息整合模型如图 5.14 所示。

3. 信息到达阶段

在信息的到达阶段实现“前置机→业务库”的对接，解决信息流转过程中的“最后一公里”问题。

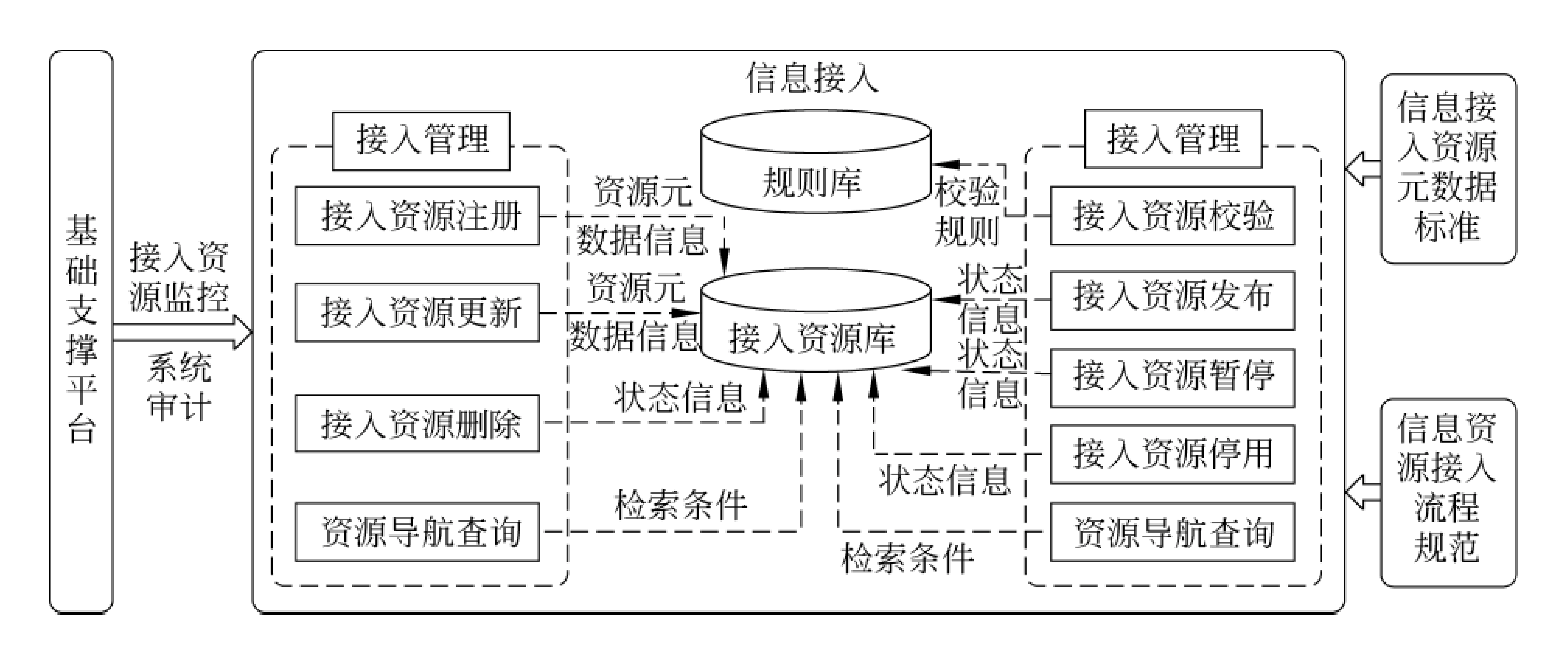


图 5.13　城市系统下的信息接入模型

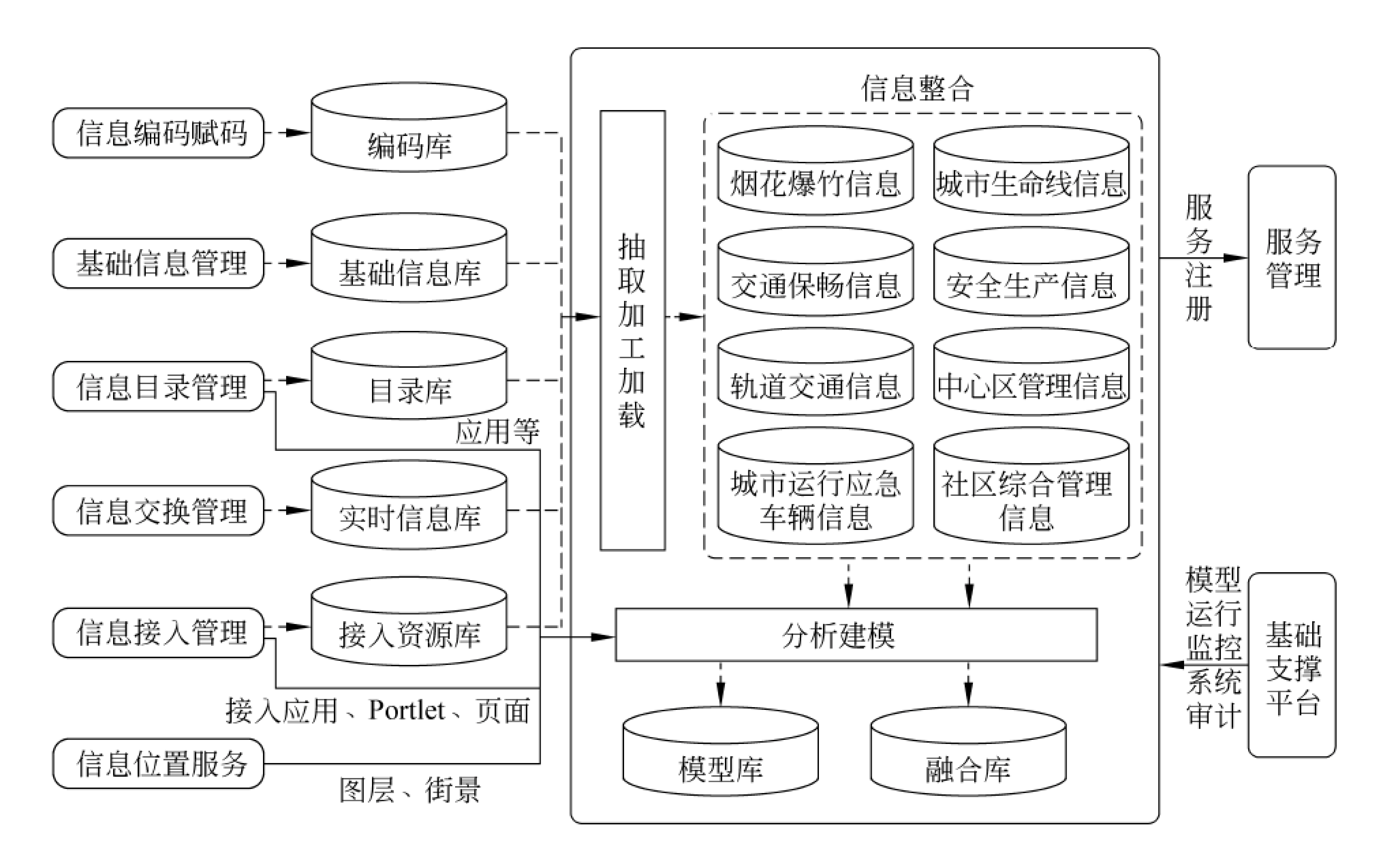


图 5.14　城市系统下的信息整合模型

5.3.2　多源信息协同的自适应进程

多源信息协同的自适应优化是体现城市系统下信息智慧流转的核心环节。在智慧城市信息化的顶层设计中，信息流转的自适应优化进程主要分为三个阶段(见表 5.8)。

表 5.8　信息流转的自适应阶段

阶段	自适应进程	构建重点	对接目标
1	业务流程的自适应匹配	业务流程库	技术流程→业务流程
2	业务规则的自适应选择	业务规则库	业务规则→业务流程→技术流程
3	业务内容的自适应优化	应用场景库	自适应选择→自适应优化

1. 业务流程的自适应匹配

一方面，由传统模式下的先注册信息目录、后开展信息交换，转变为基于需求直接开展信息交换，在信息流转过程中同步生成（更新）信息目录；另一方面，通过建立技术流程与业务流程的关联关系，可以对信息流转的共享情况、需求情况、应用业务、交换频率等信息进行全流程的网状查询。

2. 业务规则的自适应选择

一方面，信息流转模式可以根据业务规则进行自适应选择，当应用需求（如信息产生的频率变快，出现新的信息源等）发生变化时，触发业务规则库同步进行调整，信息协同体系能够自动根据业务规则的变化自动进行调整，选择最合适的信息协同模式；另一方面，对技术流程的调整可以同步调整相应的业务流程和业务规则库。



3. 业务内容的自适应优化

一方面，通过自适应选择过程中对技术流程的调整，反过来调整不合理的业务规则；另一方面，通过不同应用场景下的反复学习和优化，将通用的自适应选择过程模式化，不断完善基于情景推演的快速决策场景。

多源信息协同自适应优化的三个阶段体现了信息协同智能化的不同程度，其中，业务规则的自适应选择和业务内容的自适应优化形成一个不断循环迭代的过程，如图 5.15 所示。

信息协同自适应优化的智能过程是未来发展的重点方向。业务流程的自适应匹配阶段已具备一定的业务基础和技术基础，但业务规则的自适应选择和业务内容的自适应优化两个阶段需要在智能化方面开展进一步研

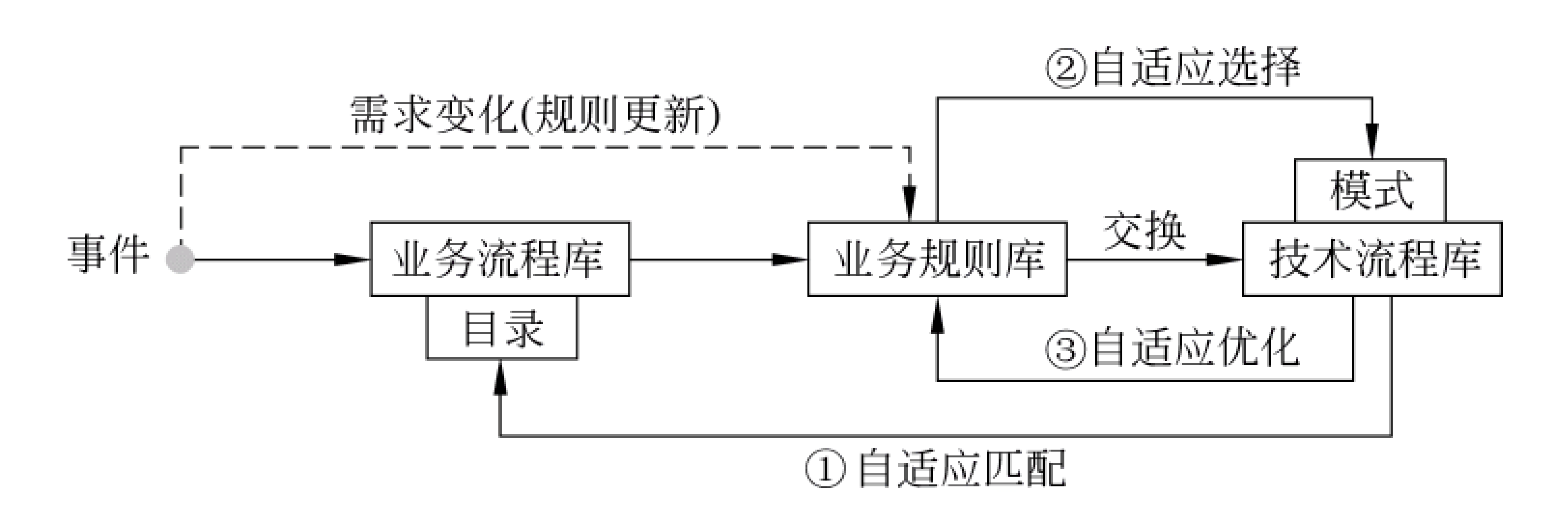


图 5.15　多源信息协同的自适应进程

究。其中,自适应选择的智能化重点在于需求变化触发规则变化后,模式自适应调整的及时性、准确性和总体资源的合理分配;自适应优化的智能化重点在于对大量场景和案例的反复变化和选择过程中,对模式和规则之间的最优化关联。

5.3.3　事件驱动的多源信息融合

基于事件驱动的多源信息融合需要构建"数据-方法-应用"三元组,其基本范式如下:

Scenario={Data, Mothod, Application}
Data={基础信息,专题信息,事件信息}
Mothod={关联规则挖掘方法}
Application={基础信息、专题信息与事件的关系}

上述范式简记为 S={D, M, A}。其中,基础信息主要包括人口基础信息、法人基础信息、空间地理基础信息、宏观经济基础信息等,专题信息包括实有房屋专题信息、气象专题信息、交通专题信息、城市生命线专题信息等。

利用上述范式实现城市管理的规律验证并进行预警,如流动人口密度或流动人口比例达到一定数值会存在较大的潜在事件发生概率,并根据概率实现预警。

多源信息融合的技术模型由数据转换引擎、数据关联引擎、数据切分引擎、数据聚合引擎和信息服务引擎组成(见图 5.16)。



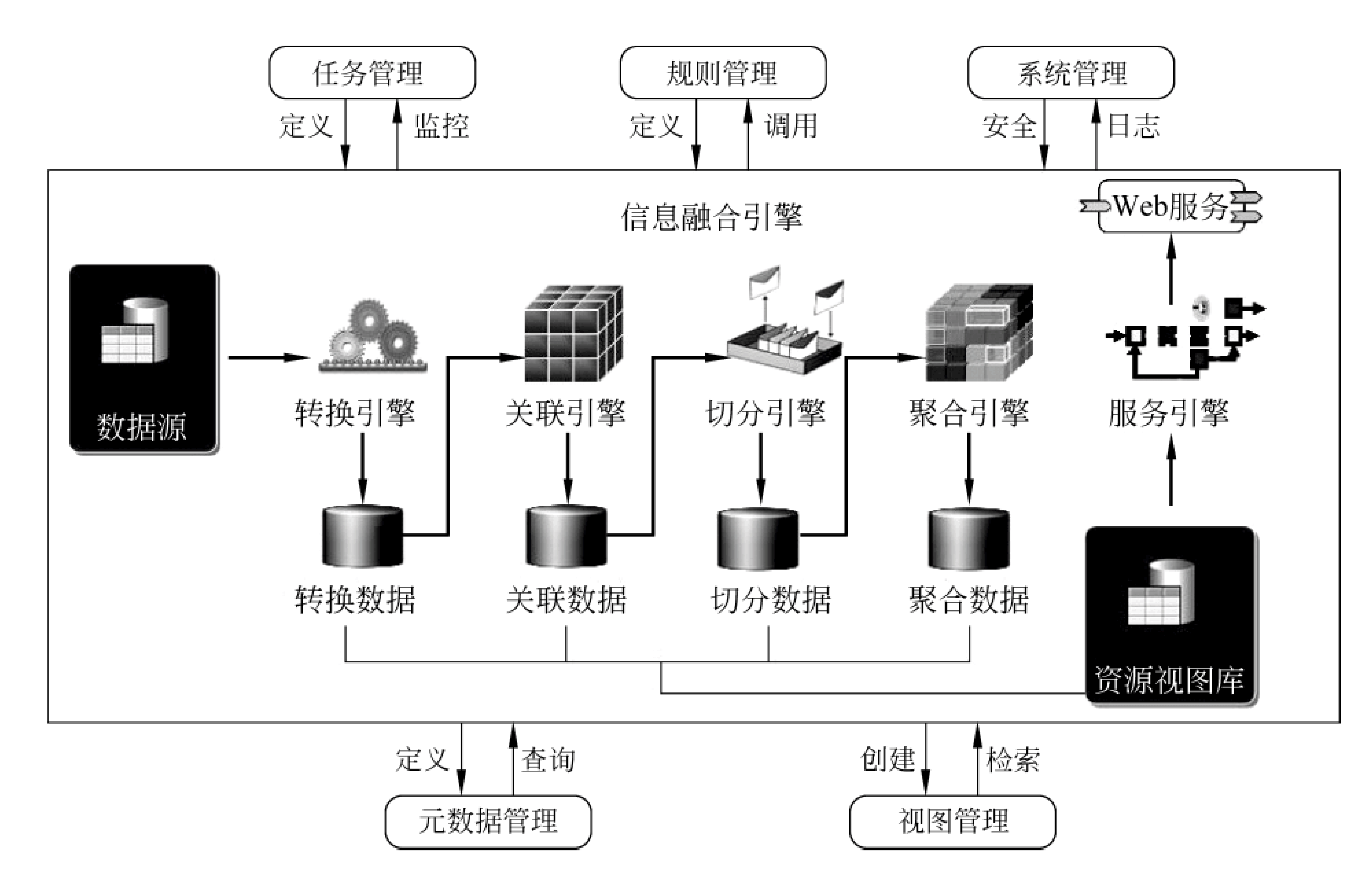


图 5.16　多源信息融合的技术模型

1. 数据转换引擎

主要对源数据进行清洗和转换。数据转换根据源数据管理，对源数据进行清洗和转换，形成数据融合引擎的基础数据库。

2. 数据关联引擎

在基础数据和主题数据之间建立各种关联关系，并形成关联数据库。这种关联包括数据库间的关联、系统间数据关联及与第三方接口的关联，并将关联处理采用关联适配器的方法进行管理，实现动态建立和维护各种关联适配器。关联适配器是按定制的关联方法对基础数据建立的关联，并形成关联数据。关联适配器管理是对各种关联适配器动态注册和注销的管理，实现动态建立和维护各种关联适配器。关联数据管理是描述各个关联适配器产生的关联数据的存储结构和结构关系，使其他部件能够正确引用关联数据。关联控制为任务管理部件提供适配器目录，根据任务管理的指令，调用相应关联适配器，并返回调用适配器的运行状态。

3. 数据切分引擎

数据切分根据切分规则，对关联数据进行切分，并形成切分数据库。切分规则管理是对各种切分方法的管理，实现动态建立和维护各种切分方法。切分数据管理描述各个切分方法产生的切分数据的存储结构和结构关系，使其他部件能够正确引用切分数据。切分控制为任务管理部件提供切分目录，根据任务管理的指令，调用相应切分方法，并返回所调用切分方法的运行状态。

4. 数据聚合引擎

数据聚合是根据聚合规则，对切分数据进行聚合，并形成聚合数据库。聚合规则管理是对各种聚合方法的管理，可实现动态建立和维护各种聚合方法。聚合数据管理描述各个聚合方法产生的聚合数据的存储结构和结构关系，使其他部件能够正确引用聚合数据。聚合控制为任务管理部件提供聚合目录，根据任务管理的指令，调用相应聚合方法，并返回所调用聚合方法的运行状态。

5. 信息服务引擎

提供服务形式的管理，主要包括信息交换、Web 服务、网页服务等多种形式。

在多源信息协同的自适应模式基础上，需要进一步构建面向城市和区域管理的多维协同机制。城市和区域的精细化管理需要面向标准、技术、应用、产业等不同方面实现从信息协同到决策行为的转化（见图 5.17）。

（1）实现组织层面的空间协同，通过区域间的横向协同、层级间的纵向协同、部门间的交叉协同实现城市和区域管理多源异构信息的纵向汇聚和横向整合。

（2）实现业务层面的领域协同，通过政党、政府、企业、公众四位一体的领域信息化在社会网格中的融合实现大数据对领域性科学决策的智慧支撑。



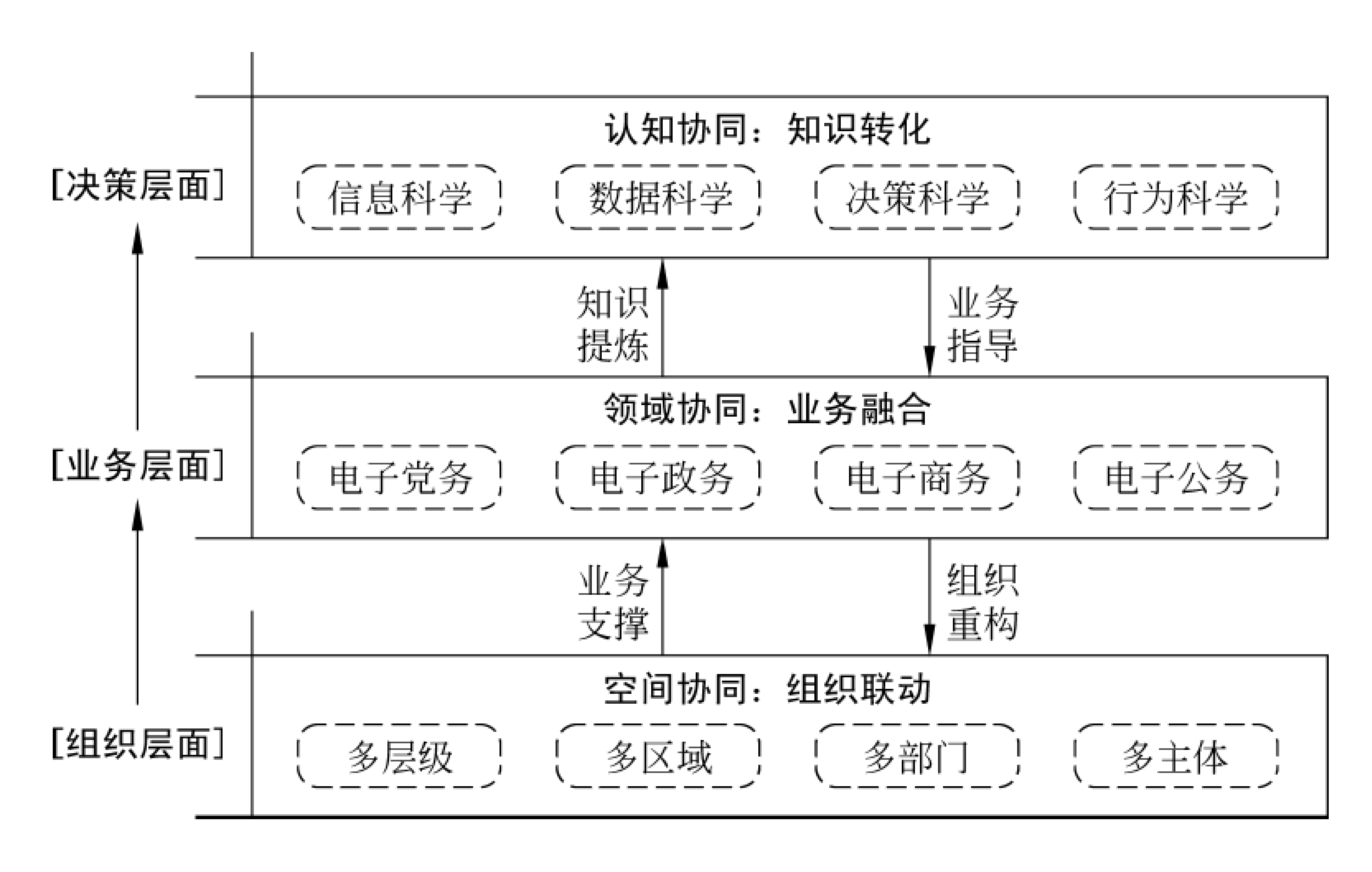


图 5.17 城市和区域管理的多维协同体系

(3) 实现决策层面的认知协同，通过物理世界和社会网络中的大数据关联分析，为解决智城市和区域管理中的空气质量、交通拥堵、水质监测、食品安全、公共安全、社区治理等突出问题提供科学支撑。

第 6 章　智慧城市多源信息协同的总体架构

本章重点介绍智慧城市多源信息协同的总体架构(包括各部分间的接口)及管理服务框架,对数据传输、数据抽取、数据转换等传统技术和建模过程不进行具体描述。

6.1　多源信息协同体系的技术架构

1. 多源信息协同体系的层次结构

多源信息协同体系的层次结构由信息协同管理中心、信息协同业务中心和信息协同节点三层构成,见图 6.1。

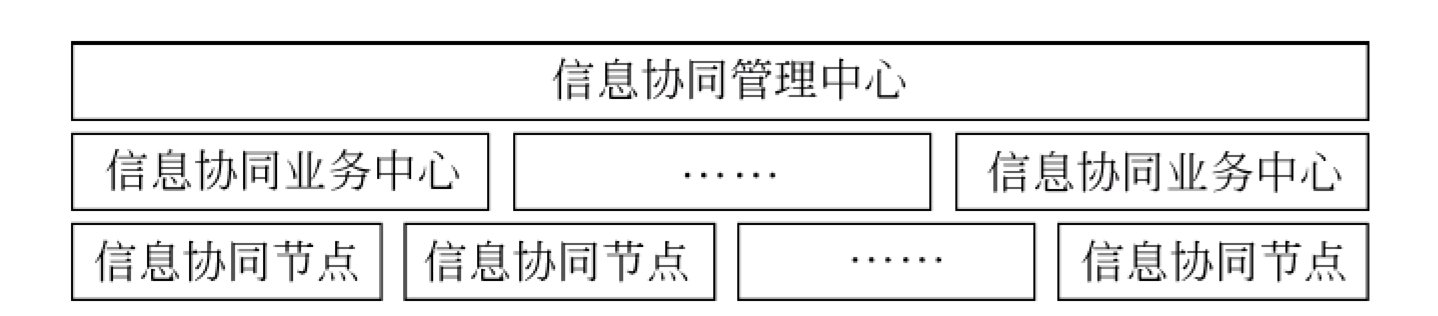


图 6.1　多源信息协同体系的层次结构

(1) 信息协同管理中心:信息协同总体架构的总枢纽和总调度,部署在中心端,提供对可配置的信息交换、处理与整合服务,及对各种应用的综合管理和服务集成,并负责信息协同总体架构内的统一授权、资源分配、运行监控和管理。

(2) 信息协同业务中心(领域中心):信息协同总体架构的中间层,部署在业务负责(领域牵头)部门,负责该领域的信息协同业务规则、业务模式的制定,信息交换流程的设计、配置、运行监控和管理,并负责对领域内各个信息协同节点的授权管理。



（3）信息协同节点：信息协同总体架构的前置端，部署在各部门，负责实现与相关业务中心之间的信息交换。

2. 多源信息协同体系的技术架构

多源信息协同体系的技术架构主要包括中心服务总线、中心管理服务、数据应用集成、协同服务组件和协同配置服务五个部分。

（1）中心服务总线：提供集中管理的中心和节点协同消息服务。

（2）中心管理服务：提供信息协同节点之间和节点内部（桥接服务）端到端的数据交换与整合服务。

（3）数据应用集成：提供连接到中心管理服务的信息协同流程设计、部署与监控服务。

（4）协同服务组件：提供通用协同服务组件的授权共享和调用服务，包括数据库适配器、转换组件、基于内容的路由组件、Web Service、REST 服务组件等，协同服务组件支持 JMS、JCA 组件标准。



（5）协同配置服务：提供中心服务总线的可视化接口，同时为信息协同中心和信息协同节点提供全局时间同步服务。

除此之外，中心管理服务和其他外部通信服务实现对接，保证对事件及异常消息的及时传输和响应。

多源信息协同体系的技术架构如图 6.2 所示。

3. 多源信息协同体系的通信与数据交换模式

多源信息协同体系架构中主要存在管理监控通信、数据交换通信、事件通信、日志通信、基本信息同步和时间同步六类通信。

（1）管理监控通信：基于 JMX 管理框架，通信协议包括 RMI、HTTP 等多种类型。

（2）数据交换通信：支持消息中间件、FTP、HTTP 等多种方式。

（3）事件通信：信息协同节点与信息协同中心之间的各类事件监控通信。

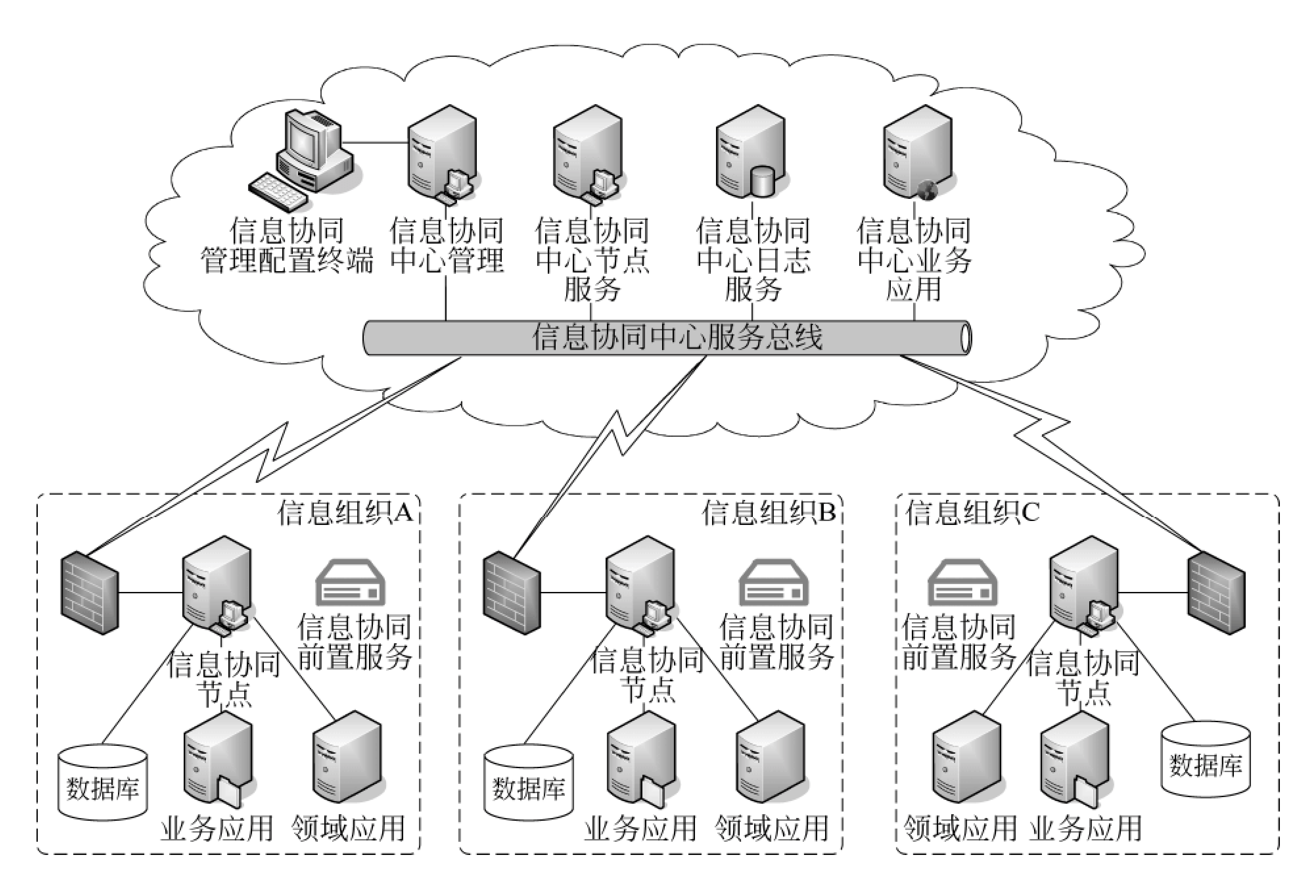


图 6.2 多源信息协同体系的技术架构

(4) 日志通信：信息协同流程的日志汇总。

(5) 基本信息同步：管理中心与业务中心之间的基本信息同步服务。

(6) 时间同步：通过时钟同步协议实现信息协同中心和节点的统一授时。

数据交换是多源信息协同体系架构的核心，对应不同业务层次的需求，主要包括以下三种数据交换模式：

(1) 批量、增量单向交换模式：一对一或一对多模式，可以是定时或实时、全集覆盖或增量交换，采用数据增量触发、时间规则触发的事件驱动架构(EDA)异步消息方式的交换技术，应用于数据交换和整合。

(2) 请求/响应服务模式：数据分布在各应用领域内，通过调用服务接口获取数据，应用于应用服务整合。

(3) 订阅/发布模式：信息提供方根据订阅规则向信息需求方进行数据推送。由于采用广播方式，因此不适用于实时信息协同场景。

6.2 多源信息协同体系的功能架构

1. 中心服务总线

中心服务总线对多源信息协同总体架构进行集中控制，可以作为其他所有信息协同节点的监控代理，并确保信息一致。主要功能包括三个方面：一是在信息协同节点上作为事件流程的一部分来控制信息协同组件的启动和终止；二是保持所有信息协同节点、业务组件和事件流程的更新状态；三是构建业务组件的双节点服务，实现信息协同主节点与从节点的实时互备。

多源信息协同总体架构采用分布式的组件执行方式，每一个组件（执行工作流的一部分）和集结组件的端点服务需要通过中心服务总线进行远程配置，主要包括：

（1）在多源信息协同总体架构中建立连通各节点服务的事件流程。

（2）执行存储功能，保持相关业务组件及其中的资源和数据的历史版本。



（3）以 XML 格式表示事件流中的元数据信息。

（4）表现及可用性管理：保持信息协同网络中所有信息协同节点的状态信息。

（5）事件跟踪、监视和调试：保持信息协同网络中所有信息协同节点的监视信息、日志信息和状态信息。

（6）安全控制：包括网络/协议层的安全和用户层的安全。协议层安全需要支持 HTTP 和 SSL 协议，用户层安全主要防止通过节点传入错误或受损的数据。新客户端连接到信息协同网络中需要提供信任并通过审计，中心服务总线通过下层服务（不同的存储与认证机制）来完成用户身份的认证并完成相应安全策略。这一安全架构允许为每个可能在信息协同网络中执行的操作进行访问控制，例如对每个事件流建立访问控制列表来识别哪些用户拥有进入网络的权限，对每个业务组件建立访问控制列表来识别该业

务组件在网络中得以运行的地点。

(7) 失效转移管理：当一个业务组件的主节点运行崩溃时，在从节点(备份节点)上配置一个新的组件实例。

2. 中心管理服务

中心管理服务为多源信息协同总体架构提供管理协同信息、监控协同状态的集成服务，方便对信息协同总体运行情况的掌控，及时发现信息协同过程中的问题，准确、快速地定位问题原因，保证信息协同流程的正常与稳定。其主要功能包括信息协同的流程管理、服务管理、目录管理、日志管理、监控管理、审计管理、权限管理、字典管理等。

(1) 多源信息协同流程管理：一是信息协同的技术流程管理，提供信息协同技术流程的配置、监控和管理服务，其中监控服务主要包括对信息协同流程运行状态的信息记录、信息查看、故障和报警提示等，管理服务主要包括信息协同流程信息(流程名称、运行状态、发送方、接收方等)的注册、启动、修改、停止、注销等；二是信息协同的业务流程管理，提供信息协同业务流程及其与技术流程关联关系的管理。

(2) 多源信息协同服务管理：提供 Web 服务、JMS 服务、REST 服务等服务内容(服务地址、分类、状态等)的注册和管理。

(3) 多源信息协同目录管理：提供对协同信息的目录管理和元数据描述。

(4) 多源信息协同日志管理：提供信息协同流程的运行日志和数据交换日志等信息的管理。其中，运行日志包括流程交换日志、节点日志(节点名称、节点代码、节点状态、启停时间)、技术流程日志(流程名称、流程代码、流程状态、启停时间)、业务流程日志、业务数据跟踪日志(业务数据在某一业务流程内的流向及最终位置)、报警日志等；数据交换日志包括各类数据的历史交换情况和交换轨迹。

(5) 多源信息协同监控管理：提供管理中心、信息协同节点、信息协同



流程和服务组件运行情况监控和管理,并根据监控结果进行多维统计分析。

(6) 多源信息协同审计管理:提供对审计内容的规则定制和配置,基于需求进行审计。

(7) 多源信息协同权限管理:提供对信息协同总体架构中的用户、权限、角色等的定义、授权和管理。

(8) 信息协同字典管理:提供对信息协同相关各类信息的定义和管理。

3. 数据应用集成

数据应用集成服务运行在业务(领域)中心和信息协同节点端,与中心服务总线和中心管理服务进行对接和实时交互,在不改变信息协同节点原有业务流程的前提下,与中心端共同完成信息协同的流程配置(流程设计、接口设置、参数配置、流程部署等)、流程管理(流程启停、流程监控等)和数据处理(数据适配、数据转换、数据加密/解密、数据整合、数据加工等)工作。

4. 配置和管理



配置和管理功能主要包括信息协同流程与中心服务总线的对接(访问保存的进程和运行的流程)和数据映射(定义事件流程中的数据转换)、服务和安全管理、事件管理(查看正在执行的事件流程运行信息)、部署管理(运行部署规则来控制组件的开发和执行)、网络管理等。

5. 高可用性设计

对复杂巨系统而言,要满足业务实时性和不间断性上的需求,就需要信息协同总体架构在消息、中心服务和节点服务三个级别上具备大数据量下的高可用性(High Availability,HA)。信息协同总体架构建立在主代理服务(Active)和从代理服务(Passive)的基础上,通过反向通道同步和容错服务连接,主从服务之间通过独立的 HA 通道实时同步,如图 6.3 所示。

后台通道同步的核心是主服务将其存储的数据和状态复制到从服务中,保持了主从服务的同步。同步通道专用于同步代理状态和消息数据的专有网络。主从服务使用同步通道实时监测其他服务的数据流程和连接,

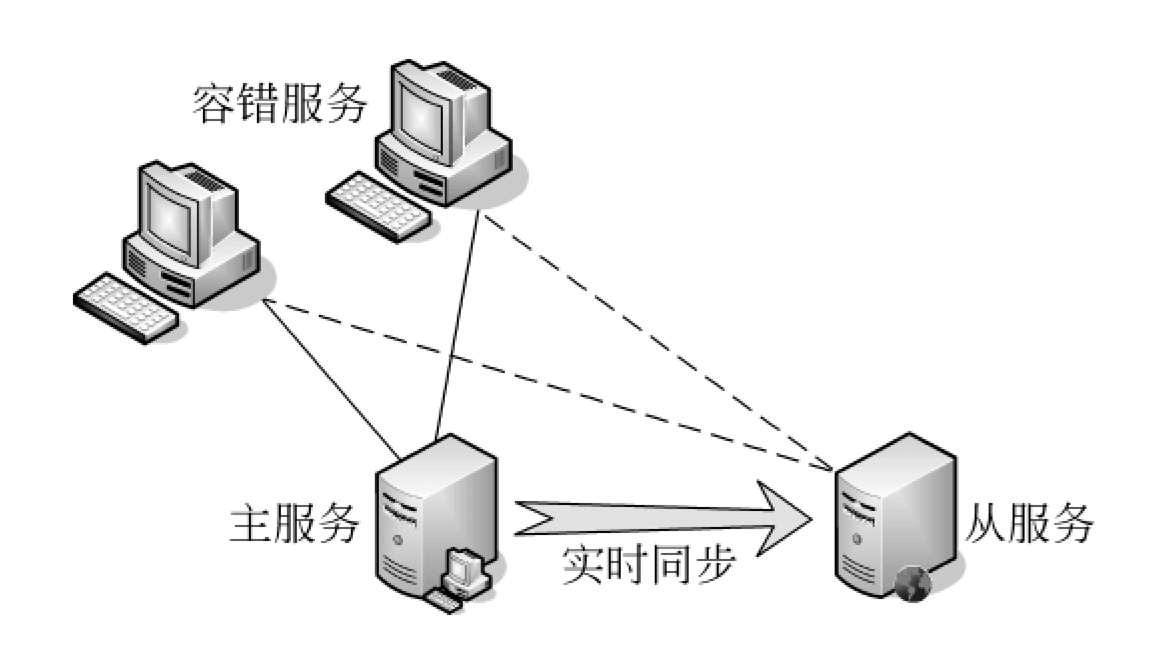


图 6.3 主从服务间的 HA 通道同步

当次服务作为热备作用时不接受节点连接，但是一旦其探测出主服务不可用则即时切换为主服务状态，所有主服务中的流程和数据同步切换到对应的从服务上。以动态作用的从服务对同步通道的重新建立（主服务恢复）进行实时监测，一旦主服务启动，则实时切换到从服务状态。

中心服务和节点服务的高可用性架构重点在于主从服务间的故障切换配置（见图 6.4）。

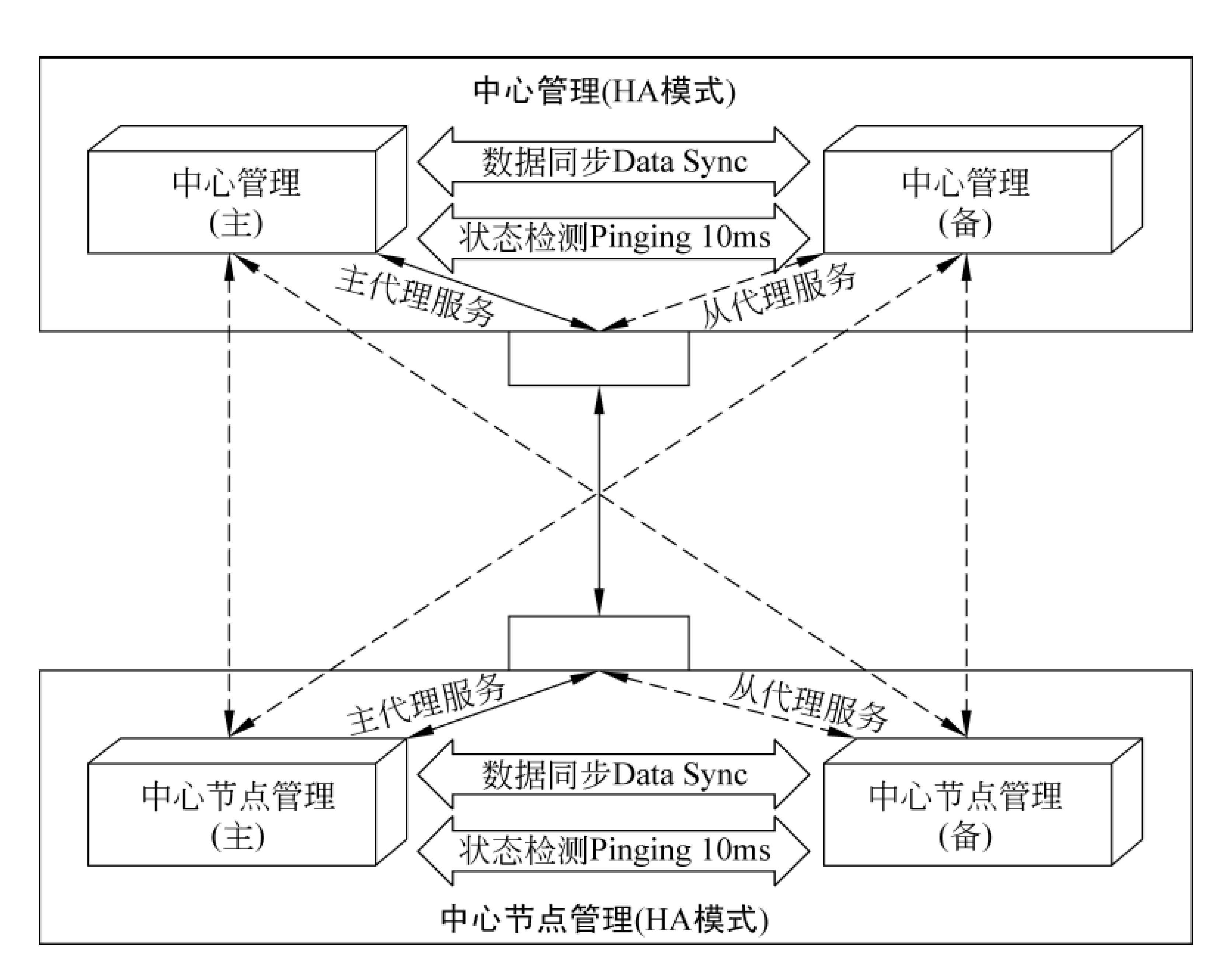


图 6.4 中心服务和节点服务的 HA 架构

6.3 数据分布和接口架构

6.3.1 数据分布架构

多源信息协同的数据分布架构如图 6.5 所示，其中信息均为信息协同流程的运行信息，不包括业务数据的描述。

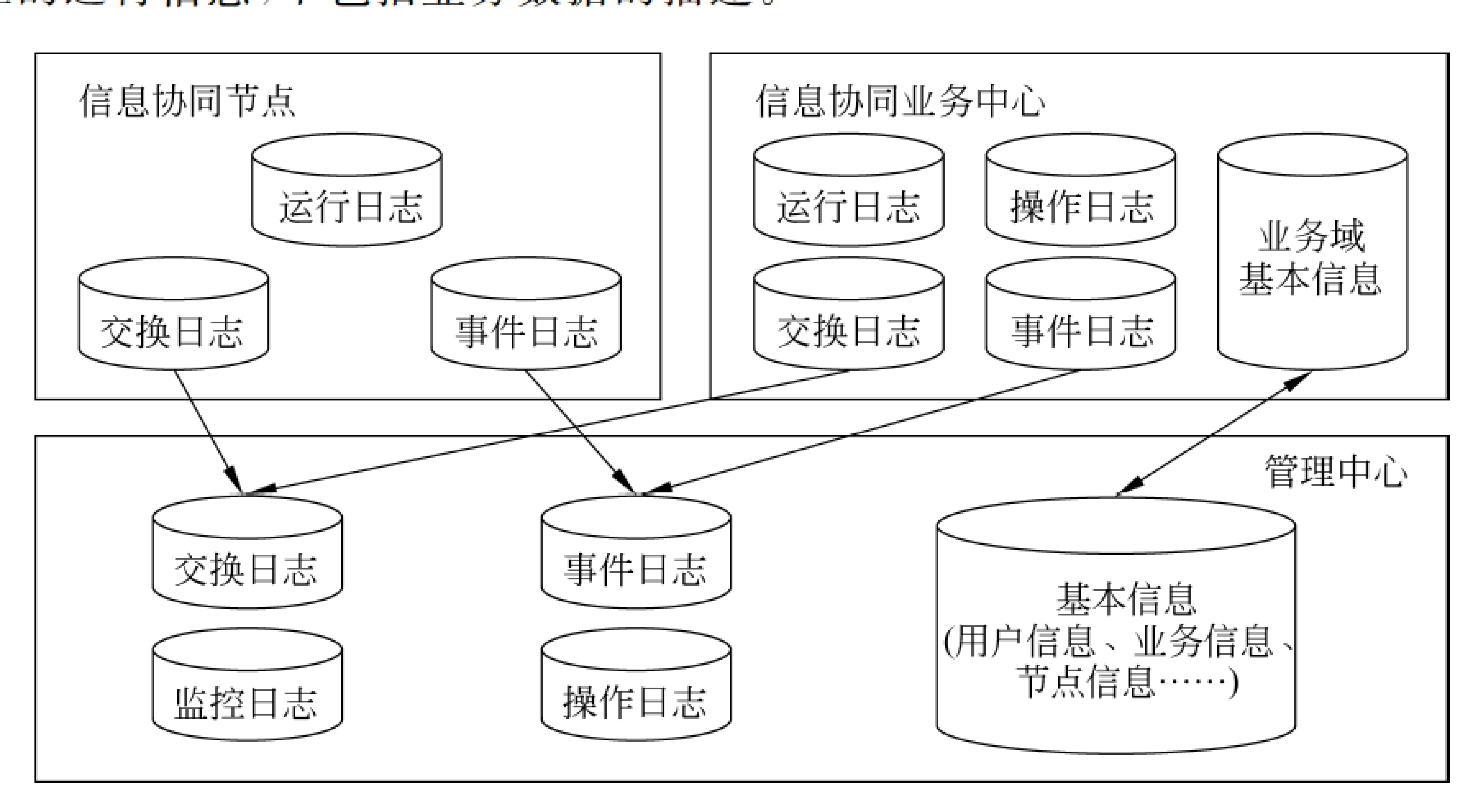


图 6.5　多源信息协同的数据分布架构



1. 信息协同节点的信息

在信息协同节点上主要包括三类信息。

（1）运行日志信息：指信息协同节点管理器、信息协同中间件运行记录的日志信息。

（2）事件日志信息：对于发生的需要被管理监控的系统行为进行记录，形成本地的事件日志，事件在发生时即时通知管理中心的事件管理服务。

（3）交换日志信息：在数据处理流程运行过程中，对数据交换的情况的记录。

2. 信息协同业务中心的信息

在信息协同业务中心主要包括五类信息。

（1）运行日志信息：采用直接连接到信息协同节点提供的 Web 日志浏览的方式。

（2）事件日志：与信息协同节点上的时间日志信息保持同步。

（3）交换日志信息：与信息协同节点上的时间日志信息保持同步。

（4）操作日志信息：根据审计策略定义，对通过信息协同业务中心进行管理操作的记录，提供查询统计功能。

（5）业务域基本信息：实现和管理中心的基本信息的同步，实现信息协同业务中心在管理中心无法联通情况下的独立运行，主要包括用户信息（基本信息、角色信息、授权信息等）、注册信息（业务注册信息、节点注册信息）、审计策略定义信息等。

3. 管理中心的信息

在管理中心主要包括五类信息：

（1）交换日志信息：对来自于各信息协同节点交换日志信息，按照统一的数据要求标准进行汇总并提供查询和统计服务。

（2）事件日志信息：接收来自信息协同节点的节点管理器、信息协同中间件运行过程中各类事件信息的汇总，这些事件发生后需及时通知管理中心的事件通知服务，由事件管理服务处理并保存，包括异常事件类型、节点号、异常事件描述、异常处理办法、异常节点是否已短信等方式通知、异常发生时间等。

（3）监控日志信息：管理中心监控服务采用定时轮询的方式对信息协同总体情况进行监控，监控日志是对每次监控轮询中，每个被监控对象的运行状态的客观记录。

（4）操作日志信息：根据审计策略定义，对通过管理中心进行管理操作的记录，包括操作类型、操作结果、操作人、操作人机器 IP 地址、操作时间等。

（5）基本信息：包括用户信息（基本信息、角色信息、授权信息）、业务域注册信息（标识支撑业务的信息）、节点注册信息（标识节点、节点管理器、



信息协同中间件等的注册信息)、监控策略定义信息、审计策略定义信息等。

6.3.2 接口设计

在多源信息协同的用户接口设计上，管理中心服务和信息协同业务中心服务采用B/S模式；数据交换的建模及配置过程采用C/S模式，提供建模、部署及管理监控功能。

多源信息协同的功能接口的设计主要包括节点远程管理接口、信息协同流程本地管理接口、信息协同流程远程管理接口、日志接口和事件接口五类，具体接口和功能描述如表6.1所示。



表6.1 多源信息协同的功能接口

序号	接口分类	接口名称	功能描述
1	节点远程管理接口	信息协同流程启动、停止	管理信息协同流程的启动、停止
2		信息协同流程状态浏览	获取信息协同流程当前的运行状态
3	信息协同流程本地管理接口	信息协同流程启、停接口	管理信息协同流程的启动、停止
4		信息协同流程运行状态获取	获取信息协同流程当前的运行状态
5		信息协同节点运行情况获取	获取节点的CPU、内存等运行情况
6	信息协同流程远程管理接口	数据处理流程查询接口	支持全部及带条件的流程查询
7		数据处理流程调度配置接口	对数据流程的调度
8		数据处理流程属性读、写接口	对数据处理流程的属性读取和修改
9		数据处理流程运行状态读取	获取数据处理流程的运行状态
10		连接器查询接口	支持全部及带条件的查询
11		连接器属性读、写接口	对连接属性读取和修改
12		传输通道(监听服务)查询	对数据传输/监听服务的查询
13		传输通道配置读取接口	读取传输通道的配置信息
14		传输通道配置修改接口	修改传输通道的配置信息
15		JMS传输队列增、删、改接口	实现对JMS队列的动态配置
16		JMS队列消息监控接口	读取JMS队列中待处理的消息数量
17		监听服务器启、停接口	对HTTP/s、Web Service等方式的启停
18	日志接口	交换日志汇总接口	将分布在各节点的交换日志汇总到管理中心
19	事件接口	事件通知接口	在管理中心提供事件监听服务和端点认证机制，实现全网事件记录

下篇 内生结构：

多源信息协同网络的测度与优化

协同主体差异和协同主体关系是反映一个协同网络**内生结构**的两个必要元素，二者相辅相成，共同对网络的内部世界进行画像。差异测度利用信息组织的“**属性变量**”，体现协同结构中信息组织的特征相似程度。关系测度利用信息组织间的“**关系变量**”，体现协同网络中存在的凝聚子群及其相互关系。

第7章　基于模糊聚类的协同网络差异测度



不同主体间的多源信息协同是一个复杂巨系统正常运转的前提。传统的城市系统下的多源信息协同模式以中心管理模式为主，协同结构相对单一，在智慧城市复杂巨系统的应用中，协同效率比较低下。近年来，随着4G、移动智能终端、物联网等新一代信息技术应用的爆发式发展，对汇聚多源信息的城市大数据应用提出了更高的要求，这也进一步加剧了信息协同效率提升与协同结构相对单一的矛盾。如何对智慧城市的多源信息协同网络进行科学测度，并在此基础上对协同模式进行合理优化，已成为制约智慧城市深化发展的关键科学问题。

目前，对智慧城市多源信息协同网络的测度主要集中在总体层面，如协同主体的规模和覆盖率、信息编目和信息共享的数量等，缺少对内生结构的测度。在本章和下章中，分别对两种维度的结构测度方法及相应优化策略进行介绍。

一是对协同主体的差异测度及横向优化。聚类分析根据对象的特征（属性）大小分类，体现了对象个体特征（属性）的相似程度。针对信息组织和信息领域的交叉性和关联性特征，模糊聚类在特征相似度分析的同时可以给出置信度区间，是研究复杂巨系统下多源信息协同网络的有效方法。

二是对协同主体间的关系测度及纵向优化。以聚类算法为基础的凝聚子群分析属于社会网络分析范畴，也是研究结构问题的主要方法之一。凝聚子群根据对象之间的相互关系分类，体现了对象间的关联关系，是对网络进行多层次结构分析的主要方法。

7.1　硬聚类与模糊聚类

7.1.1　聚类分析的基本概念与方法

“物以类聚，人以群分”，这句话最恰当地描述了聚类分析的目的。对小数据量的数据进行分类处理时，比如个位数的数据，可以手工对数据进行分类，但是当数据量变大时手工分析处理每个对象就变得不切实际。聚类分析的实质是建立一种预先未知的对样本或变量进行分类的技术，对样本（对象）的聚类称为 Q 型聚类，对变量的聚类称为 R 型聚类。一般情况下，主要是对样本进行聚类，从而在没有先验知识的情况下将样本自动分类。

聚类分析可以看作是一种无教师的模式分类方法，在分类时只依赖对象自身所具有的属性来区分对象之间的相似程度。当前聚类分析算法的研究中也有一些有学习过程的聚类算法，如半监督、有监督的聚类算法，但通常情况下聚类分析算法都不需要提供预先已知类别信息的样本来训练分类器。聚类分析算法作为一种有效的数据分析方法被广泛应用于数据挖掘、机器学习、图像分割、语音识别、生物信息处理、商业分析等领域；聚类算法还可以应用于商业分析，它可以帮助市场决策人员从消费者数据库中区分出不同的消费群体，并且概括出每一类消费者的消费模式或习惯。从本质来说聚类算法是将总体中的个体分类以发现数据中的结构，希望一个类中的个体彼此接近或相似，而与其他类中的个体相异，这样就可以对划分出来的每一类进行深入分析从而概括出每一类的特点。

目前，聚类算法主要分为层次化聚类方法、划分式聚类方法、基于密度的聚类方法、基于网格的聚类方法、基于核的聚类算法、基于谱的聚类方法、基于模型的聚类方法、基于遗传算法的聚类方法、基于 SVM 的聚类方法、基于神经网络的聚类方法等。

1. 聚类算法的基本定义

给定一个对象集合 $X=\{x_1,x_2,\cdots,x_n\}$，假设每个对象 $x_i(i=1,2,\cdots,n)$ 含有 m 个特征，在此用向量的方式来表示对象的特征，$x_l(l_1,l_2,\cdots,l_m)$，聚类分析的过程就是根据对象的特征来分析对象之间的相似程度，并根据某种聚类决策准则来获得聚类结果。聚类的结果用 $C=\{c_1,c_2,\cdots,c_k\}$ 表示，则聚类结果满足以下条件：

（1）$c_i\neq\varnothing,i=1,\cdots,k$；

（2）$\bigcup_{i=1}^{k}c_i=X$；

（3）$c_i\cap c_j=\varnothing,i\neq j,i,j=1,\cdots,k$。

模糊聚类的结果没有上面的约束条件，模糊聚类给出的结果只是对象隶属于每个类的程度。通常聚类分析算法一般包含四个部分：①特征获取与选择；②计算相似度；③分组；④聚类结果展示。

2. 距离和相似系数

在聚类分析中，用距离描述对象间的靠近程度，用相似系数描述变量间的联系紧密程度，距离和相似系数可以相互转换。因为不同的相似性指标所测量的相似程度的意义有所区别，因此选用不同的距离或相似系数，可能会得到不同的分类结果。

距离测度方法主要有欧几里得距离（Euclidean Distance）、明可夫斯基距离（Minkowski Distance）、曼哈顿距离（Manhattan Distance）、切比雪夫距离（Chebyshev Distance）、马哈拉诺比斯距离（Mahalanobis Distance）等；相似度测度方法主要有向量的夹角余弦（Cosine Similarity）、皮尔森相关系数（Pearson Correlation Coefficient）、Jaccard 相似系数（Jaccard Coefficient）等。欧氏距离是最常见的距离度量，向量的夹角余弦是最常见的相似度度量，很多距离度量和相似度度量都是基于这两者的变形和衍生。

距离度量衡量的是空间各点间的绝对距离，与各个点所在的位置坐标（即个体特征维度的数值）直接相关，体现个体数值特征的绝对差异，主要用

于需要从维度的数值大小中体现差异的分析，如使用用户行为指标分析用户价值的相似度或差异。向量的夹角余弦衡量的是空间向量的夹角，更多的是体现在方向上的差异，而不是位置，对绝对的数值不敏感，主要用于使用用户对内容的评分来区分用户兴趣的相似度和差异，同时修正了用户间可能存在的度量标准不统一的问题。

除以上所列出的常见距离函数和相似系数外，还有一些专门用于测量离散值向量的距离函数、混合值类型的距离函数，及专门用于模糊集的距离函数等。在聚类分析中还涉及两种类型的距离计算：一是类间距离（Inter-cluster Distance），表示两个类之间的差异；二是对象和类之间的距离，表示对象和一个类之间的差异。计算模式之间相似程度的研究是目前模式识别领域的一个重要方向。比较有代表性的研究有马氏距离函数在高维数据空间中的信息丢失问题、基于分布模型和距离函数的统计分析找到最好距离来估计对象之间相似度的通用向导、马氏类型数据的距离函数比较、生物序列距离的快速算法、无序列比对的蛋白质序列距离估计方法、用于计算 SCOP 蛋白质数据集的序列距离计算方法等。除此之外，由于距离函数的设计和聚类分析的对象密不可分，在一些新的应用领域出现了一些新的或改进的距离计算方式，如基于自适应的 Hausdorff 距离函数的动态聚类算法、基于压缩距离的蛋白质序列分类算法、基于点对称的距离测度的进化聚类算法、采用内部距离分类形体的算法、通过类间距离选择支持向量机的参数等。

3. 主要的硬聚类算法

聚类是数据库知识发现领域的重要课题。目前聚类算法主要有层次聚类方法、划分式聚类方法、基于密度的聚类方法、基于网格的聚类方法、基于核的聚类算法、基于谱的聚类方法、基于模型的聚类方法、基于遗传算法的聚类方法、基于 SVM 的聚类方法、基于神经网络的聚类方法等，各种方法有不同的基本思想。

1）层次聚类

层次聚类算法（Hierarchical Clustering）又称为树聚类算法或系统聚类算法，是最常用的硬聚类算法之一，按自底向上层次分解称为凝聚法（agglomerative）层次聚类，按自顶向下层次分解称为分裂法（divisive）层次聚类。这种算法的基本思路是首先将所有对象看成独立的个体类，通过计算类间的距离来选择最小距离的两个类合并成一个新类，再重新计算新类和其他类之间的距离，选择最小距离的两个类合并，依次迭代合并直到没有合并为止。

层次聚类过程中按照类间距离计算方式的不同主要有以下几种方法：

（1）最短距离法（单一连接、紧邻连接）：两个类之间的距离定义为两类中元素之间距离最小者。

（2）最长距离法（完全连接、最远紧邻连接）：两个类之间的距离定义为两类中元素之间距离最大者。



（3）中间距离法：两个类之间的距离定义为两类中元素之间的某个中间距离。

（4）重心法：两个类之间的距离定义为两个类的重心间的距离。

（5）类平均法：两个类之间的距离（平方）定义为两类中元素两两之间的平均（平方）距离。

（6）变差平方和法（WARD法）：与方差分析思想类似。在分类过程中，使类内元素间的变差平方和尽可能小，而类与类之间的变差平方和尽可能大。

如果聚类过程中每一步聚类时的距离都大于前一步，则称该聚类方法是单调的。如果两个类之间的距离基本取中间，既不取最短（空间收缩），也不取最长（空间扩展），则称该聚类方法空间守恒。单调性和空间守恒都是一个好的聚类方法的表现。以上六种主要层次聚类方法的综合比较如表7.1所示。

表 7.1 六种主要层次聚类方法的综合比较

方　法	空间性质	单调性	距离要求	适用形	结果唯一性
最短距离法	压缩	单调	无	条形、S形	唯一
最长距离法	扩张	单调	无	椭球形	距离表中有相同元素时，可能出现不唯一的结果
中间距离法	守恒	非单调	欧氏距离的平方		
重心法	守恒	非单调			
类平均法	守恒	单调	无		
变差平方和法	扩张	单调	欧氏距离的平方		

目前常用的层次聚类算法有：CURE、ROCK、BIRCH、Chameleon 等。层次聚类算法的改进算法也很多。Goldberger Jacob 提出一种基于 Hungarian 方法的层次聚类方法，该方法的输入只需要样本点之间的成对距离即可。经典 Hungarian 方法是求解最小加权环覆盖问题的有效方法，该方法使用 Hungarian 方法来构建基本的聚类块划分。Loewenstein Yaniv 等改进了经典的 UPGMA 方法，用于进行大规模的蛋白质序列聚类分析。算法可以在内存受限的环境下进行大数据量的聚类分析。Wang H. 等采用改进的层次聚类算法分析基因表达序列数据，Arifin Agus Zainal 等采用层次聚类算法对图像进行分割。由于层次聚类算法适合处理大型数据集，因此被广泛应用于分析蛋白质序列数据。

2）划分式聚类

划分式聚类方法的主要思想是：对于一个给定的 n 个数据对象的数据集，采用目标函数最小化的策略，初始时选择一定量的聚类中心或数据点，通过某种原则把数据划分到各个组中，每个组为一个簇。最典型的划分式聚类算法是 k-means 算法和 k-medoids 算法。这两种算法的改进算法非常多，应用也很广泛。

比较有代表性的研究有：针对 k-means 算法只能保证收敛到局部最优，从而导致聚类结果对初始代表点的选择非常敏感的问题。雷小锋等给出了一种叫做 K-MeanSCAN 的算法，算法采用 k-means 对数据进行多次预聚

类，对预聚类结果构造子簇的加权连通图，并根据连通性合并子簇；Xiong H. 等分析了数据集本身的分布与 k-means 算法的聚类结果之间的关系；Chung Kuo-Liang 等给出了一种基于对称距离测度的 k-means 算法；Huang Joshua Zhexue 等给出了一种在迭代划分过程中自动变换变量权值的 k-means 算法；Wu F. X 采用遗传加权 k-means 算法来分析基因表达数据；Bagirov Adil M 提出了一种新的全局 k-means 算法，算法能够克服 k-means 算法对初始点选择敏感的问题；针对 k-means 算法的类个数选择的问题，Hamerly G 等讨论了如何从聚类过程中学习 k 值的方法；Du 等将粒子群优化方法和 k-means 算法结合用于分析基因表达数据。

3）基于密度和网格的聚类

基于密度和网格的聚类算法在以空间信息处理为代表的众多领域有着广泛应用，特别是随着大规模数据集聚类的需求越来越强烈，常规的聚类算法在大规模数据集聚类分析时受到限制，因此基于密度和网格的聚类方法在空间数据挖掘研究领域日趋活跃。具有代表性的基于密度的聚类算法有 DBSCAN、GDBSCAN、OPTICS、DENCLUE、CLIQUE 等。基于网格的聚类算法常常与其他方法相结合，特别是与基于密度的聚类方法相结合。STING 是代表性的基于网格的聚类算法。



4）其他聚类算法

除了以上描述的常见聚类算法外，目前热点研究的聚类算法还有基于支持向量机的聚类算法、基于核方法的聚类算法、基于神经网络的自组织映射（SOM）算法、K-Narest-Neighbor（K-NN）聚类算法及其改进算法、基于神经气的聚类算法、谱聚类算法、复杂网络聚类方法等。目前，关于谱聚类算法的研究在图像分割、不规则形状聚类方面有很好的效果。

上面给出的聚类算法绝大多数都属于无监督的聚类算法，但在实际分析数据时，有时也能够获得一部分数据先验知识，例如部分数据的正确类信息或数据分布的信息等。利用这些先验知识来指导聚类分析，就形成了有监督或半监督的聚类分析方法。Al-Harbi SH 等给出了一种有监督的自适

应的 k-means 算法，Staiano A 等采用有监督的 FCM 方法来改进径向基神经网络的性能。利用少量的先验知识来对大量没有先验知识的数据进行聚类分析是半监督聚类分析算法的主要目的。

聚类分析算法本质上都有一个自己的分类标准，也可以理解为对数据分组的目标是什么，用数学意义上的概念来说就是目标函数。现存的大部分聚类标准或是目标函数可以归结为以下几类：

（1）以紧密性为目标的聚类分析，即簇内对象联系紧密，簇间关系疏远。代表性的算法有划分式方法中的 k-means 算法、层次聚类算法等，这类算法对球形分布的数据或者是本身可分性就很强的数据有很好的聚类效果，但是对复杂结构的数据或分布无能为力。

（2）以连通性为目标的聚类分析。这类算法的基本思路是相互邻接的数据应该有相同的模式。代表性的算法有基于密度的聚类、基于路径的聚类等。

（3）以空间隔离为标准进行聚类分析。单纯以空间隔离性作为标准难以获得很好的有意义的聚类分析结果，通常和其他标准相结合。



7.1.2 模糊聚类的基础理论

前面介绍的聚类方法可以看成是硬聚类（hard clustering），即每个对象属于并且仅仅属于一个簇，因此每个簇之间没有交集。模糊聚类（fuzzy clustering）和硬聚类不同，它采用一个隶属函数来关联对象和簇之间的关系。

一般的模糊聚类过程如下：模糊聚类算法最开始先初始化构建一个初步的划分，将数据划分为 K 个模糊组，构建一个隶属矩阵 $\boldsymbol{U}$。通过隶属矩阵求解每个模糊组的中心点，根据计算出的中心点来获得当前划分的目标函数值。将当前获得的目标函数值与上一次获得的目标函数值进行比较，如果满足截止条件则终止算法，否则更新隶属矩阵 $\boldsymbol{U}$，重复以上步骤。

Baraldi A. 等综述了模糊聚类算法，上面描述的模糊聚类过程即为

Fuzzy Mean(FCM)算法。该算法 1974 年由 Bezdek JC 提出。FCM 是目前应用广泛的模糊聚类算法之一,收敛速度快,局部搜索能力强,但对初值和噪声较为敏感,容易陷入局部最优,而得不到全局最优解。局部最优的问题一直都是困扰聚类算法的一个难题。在 FCM 算法的基础上,很多学者提出了一些改进的算法,这些改进集中在隶属函数设计、加速聚类过程、避免陷入局部最优等方面。由于原始的 FCM 算法是基于欧式距离的,即要求数据空间为球形空间,不能处理椭圆形的数据空间,王丽娟等针对这一问题提出了一种给每个特征属性加权的模糊聚类算法(CF-WFCM 算法)。由前面 FCM 算法过程的描述可知,在迭代计算的过程中要存储计算整个隶属矩阵 ***U***,并且要不断更新计算中心点,因此 FCM 算法的计算复杂度较高,难以用于大型数据集。Hathaway Richard J. 等给出了一个扩展快速 FCM 算法 geFFCM。同样为了加速 FCM 算法的运行效率,Kolen JF 等将原始的 FCM 中的交替更新隶属矩阵中耗费内存空间的过程移除,将两步更新合并为一步更新,显著加快了聚类运行效率。由于 FCM 可能过度划分数据集,Laskaris Nikolaos A. 等给出了一种 Beyond FCM 的算法,该算法增加了一个基于图的后处理阶段。Pal NR 等给出了一个中概率模糊 C 均值算法。Dembele D. 等采用 FCM 对 DNA 微阵列数据进行聚类分析;Masson Marie-Hélène 等提出了一种用于对象数据聚类的算法 ECM。

近年来,许多学者将智能化研究引入模糊聚类算法。董世龙等提出了一种基于多线程的云资源模糊聚类划分并发算法,通过传递闭包法进行优化解决高维矩阵运算问题,并将其应用于 Hadoop 调度器的策略改进;李文娟等和 Sun Da-wei 等提出了能够自适应选择的模糊聚类资源调度和分配算法;王洪斌等针对模糊建模在进行结构辨识时需事先设定聚类数的问题,提出参数自适应模糊聚类算法;Zhu L 等通过引入隶属度约束函数,提出一种改进模糊分割的聚类算法(IFP-FCM),对噪声和例外点具有更好的鲁棒性;Li Y 等给出了一种聚类数 c 自适应函数,自动给出最佳的聚类数 c 及相应的划分隶属矩阵和聚类中心,从而得到模糊辨识的前件结构和参数,即描述系

统的模糊规则和隶属度函数。随着群智能优化算法的发展，蚁群算法、微粒群算法、萤火虫算法等被引入到模糊聚类的过程中，来优化全局寻优能力和快速收敛能力。

除此之外，还有一些其他的模糊聚类算法。例如 Lee Sang Wan 等给出的迭代贝叶斯模糊聚类（Iterative Bayesian Fuzzy Clustering），Gan G. 等给出的模糊子空间聚类（Fuzzy Subspace Clustering，FSC），Gath I 等结合 FCM 和最大期望估计使算法能够有效分析簇间对象个数、密度、形状变化较大的情况，Grira Nizar 等提出一种活跃半监督模糊聚类算法，以及模糊自组织映射聚类算法、熵指数约束的模糊聚类等。

7.2 基于模糊聚类的多源信息协同差异测度模型

7.2.1 相关定义和定理

定义 1　模糊相似矩阵：若模糊关系 R 是 X 上各元素之间的模糊关系，且满足：① 自反性：$R(x,x)=1$，② 对称性：$R(x,y)=R(y,x)$；则称模糊关系 R 是 X 上的一个模糊相似关系。当论域 $X=\{x_1,x_2,\cdots,x_n\}$ 为有限时，X 上的一个模糊相似关系 $\boldsymbol{R}$ 就是模糊相似矩阵，即 $\boldsymbol{R}$ 满足：① 自反性 $I\leqslant\boldsymbol{R}(\Leftrightarrow r_{ii}=1)$，②对称性：$\boldsymbol{R}^{\mathrm{T}}=\boldsymbol{R}(\Leftrightarrow r_{ij}=r_{ji})$。

定义 2　模糊等价矩阵：当 $X=\{x_1,x_2,\cdots,x_n\}$ 为有限论域时，X 上的模糊等价关系 $\boldsymbol{R}$ 是一个矩阵（称为模糊等价矩阵）。它满足三个条件：①自反性：$r_{ii}=1$，②对称性：$r_{ij}=r_{ji}$，③传递性：$\boldsymbol{R}\circ\boldsymbol{R}\subset\boldsymbol{R}$；即 $\bigvee_{k=1}^{n}(r_{ik}\wedge r_{kj})\leqslant r_{ij}$，$i,j=1,2,\cdots,n$。

定义 3　模糊矩阵的 λ-截矩阵：设 $\boldsymbol{A}=(a_{ij})_{m\times n'}$ 对任意 $\lambda\in[0,1]$，称 $\boldsymbol{A}_{\boldsymbol{\lambda}}=(a_{ij}{}^{(\lambda)})_{m\times n'}$ 为模糊矩阵 $\boldsymbol{A}$ 的 λ-截矩阵。其中，当 $a_{ij}\geqslant\lambda$ 时，$a_{ij}{}^{(\lambda)}=1$；当 $a_{ij}<\lambda$ 时，$a_{ij}{}^{(\lambda)}=0$。显然，$\boldsymbol{A}$ 的 λ-截矩阵为布尔矩阵。

通常模糊关系不一定具有传递性，因此不是模糊等价关系，需要通过某

种方法对模糊关系进行改造。

定义 4　模糊传递闭包：设 $R\in\varphi(X\times X)$，称 $t(R)$ 为 R 的传递闭包。如果 $t(R)$ 满足：①传递性：$(t(R))^2\subset t(R)$，②包容性：$R\subset t(R)$，③最小性：若 R' 是 X 上的模糊传递关系，且 $R\subset R'\Rightarrow t(R)\subset R'$，即 R 的传递闭包 $t(R)$ 是包含 R 的最小的传递关系。

定义 5　模糊等价闭包：设 $R\in\varphi(X\times X)$，称 $e(R)$ 为 R 的等价闭包，如果 $e(R)$ 满足：①等价性：$e(R)$ 是 X 上的模糊等价关系，②包容性：$R\subset e(R)$，③最小性：若 R' 是 X 上的模糊等价关系，且 $R\subset R'\Rightarrow e(R)\subset R'$。显然，$R$ 的等价闭包是包含 R 的最小的等价关系。

定理 1　设 $R\in F(X\times X)$ 是相似关系（即 R 是自反、对称模糊关系），则 $e(R)=t(R)$，即模糊相似关系的传递闭包就是它的等价闭包。

7.2.2　信息组织的特征分析



一般情况下，信息组织的信息协同特征分为基础特征和扩展特征两部分。基础特征主要包括以下三个层面：

（1）信息组织层面：信息需求方数量、信息提供方数量等。

（2）信息内容层面：提供信息的情况、获取信息的情况等。

（3）信息协同多样性层面：信息载体的多样性、信息频率（周期）的多样性等。

对于有较大相关性的特征可合并降维处理，如信息的关联事件和应用领域与信息需求方合并为同类因素。

扩展特征主要包括以下两种情况：

（1）与信息协同水平无直接相关性的特征：如所属管理对象和感知设备的种类（或数量）。

（2）变化性强、不易定量的特征：如信息协同的重要性与具体业务、事件场景、应急态势等有较强的关联关系。

扩展特征在不同的情景下是反映信息组织的重要因素，可根据实际情

况作为基础特征的补充。

7.2.3 信息组织的模糊聚类模型

1. 特征提取与信息协同标准矩阵构建

设有 n 个信息组织（对象），每个信息组织有 m 个特征（属性）。令 $A=\{a_1,a_2,\cdots,a_n\}$ 表示信息组织的集合，$x_{ij}(i=1,2,\cdots,n;j=1,2,\cdots,m)$ 表示第 i 个信息组织的第 j 个特征，则信息协同原始矩阵 $\boldsymbol{P}$ 表示为

$$\boldsymbol{P}=\begin{vmatrix} x_{11}, & x_{12}, & \cdots, & x_{1m} \\ x_{21}, & x_{22}, & \cdots, & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{n1}, & x_{n2}, & \cdots, & x_{nm} \end{vmatrix}$$

若存在某个特征（矩阵 $\boldsymbol{P}$ 中的某列），各信息组织的该特征值相差极大，或由于某种特殊原因某个信息组织的特征值特别突出，该特征在聚类过程中的作用将被不适当地夸大。因此，为了避免特殊值的影响，首先通过公式（7.1）对原始数据进行预处理：

$$y_{ij}=\frac{x_{ij}-\bar{x}_j}{x_j^{\max}-\bar{x}_j}(1.00-M)+M \tag{7.1}$$

其中，$\bar{x}_j=\frac{1}{n}\sum_{i=1}^{n}x_{ij}$，$M\in[0.5,0.75]$。

经过对原始数据的特殊值处理后，对矩阵 $\boldsymbol{P}$ 进行标准化处理。常用的方法有线性变换、平移-极差（标准 0-1）变换、平移-标准差变换、最优值为给定区间时的变换、向量规范化等，可根据不同情况选择合适的方法或部分方法的组合。

1）线性变换

对于效益型的属性 j：

$$y_{ij}=\frac{x_{ij}}{x_j^{\max}} \tag{7.2}$$



对于成本型的属性 j：

$$y_{ij}=1-\frac{x_{ij}}{x_j^{\max}} \tag{7.3}$$

2）平移-极差（标准 0-1）变换

对于效益型的属性 j：

$$y_{ij}=\frac{x_{ij}-x_j^{\min}}{x_j^{\max}-x_j^{\min}} \tag{7.4}$$

对于成本型的属性 j：

$$y_{ij}=\frac{x_j^{\max}-x_{ij}}{x_j^{\max}-x_j^{\min}} \tag{7.5}$$

3）平移-标准差变换

$$y_{ij}=\frac{x_{ij}-\bar{x}_j}{s_j}(i=1,2,\cdots,n;j=1,2,\cdots,m) \tag{7.6}$$

其中$\bar{x}_j=\frac{1}{n}\sum_{i=1}^{n}x_{ij}$，$s_j=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2}$。

4）最优值为给定区间时的变换

对于既非效益型也非成本型的属性 j，给定最优属性区间$[x_j^0,x_j^*]$，x_j'为无法容忍下限，x_j''为无法容忍上限，令

$$y_{ij}=\begin{cases}1-(x_j^0-x_{ij})/(x_j^0-x_j') & 若\ x_j'<x_{ij}<x_j^0\\ 1 & 若\ x_j^0\leqslant x_{ij}\leqslant x_j^*\\ 1-(x_{ij}-x_j^*)/(x_j''-x_j^*) & 若\ x_j''>x_{ij}>x_j^*\\ 0 & 其他\end{cases} \tag{7.7}$$

5）向量规范化

无论成本型还是效益型属性，均可进行向量规范化：

$$y_{ij}=x_{ij}\Big/\sqrt{\sum_{i=1}^{m}x_{ij}^2} \tag{7.8}$$

与其他几种变换不同的是，向量规范化后从属性值的大小上无法分别属性值的优劣，各对象的同一属性值的平方和为 1，常用于计算各对象与某种虚

拟对象（如理想点或负理想点）的欧式距离的场合。

经过标准化处理后，得到信息协同标准矩阵 $\boldsymbol{Y}$：

$$\boldsymbol{Y} = \begin{vmatrix} y_{11}, & y_{12}, & \cdots, & y_{1m} \\ y_{21}, & y_{22}, & \cdots, & y_{2m} \\ \vdots & \vdots & & \vdots \\ y_{n1}, & y_{n2}, & \cdots, & y_{nm} \end{vmatrix}$$

2. 测度选择与模糊相似矩阵构建

针对标准化矩阵，计算各信息组织间的相似程度，建立模糊相似矩阵 $\boldsymbol{R}=(r_{ij})_{n\times m}$。计算相似程度的方法主要包括距离测度和相似度两种类型。其中，距离测度越大，说明对象间相似度越小，差异越大；与距离测度相反，相似系数越小，说明对象间相似度越小，差异越大。距离测度 d 与相似度测度 r 之间可以通过公式(7.9)进行转换：

$$r_{ij} = 1 - c \cdot d(y_i, y_j) \tag{7.9}$$

其中 c 为适当选取的参数。

1）距离测度

（1）欧几里得距离（Euclidean Distance）：基于各维度特征的绝对数值，需要保证各维度指标在相同的刻度级别。

$$d(y_i, y_j) = \sqrt{\sum_{k=1}^{m} (y_{ik} - y_{jk})^2} \tag{7.10}$$

（2）明可夫斯基距离（Minkowski Distance）：明氏距离是欧氏距离的推广，是对多个距离度量公式的概括性的表述。明氏距离中 $p=2$ 时得到的距离度量即为欧式距离。

$$d(y_i, y_j) = (\sum_{k=1}^{m} |y_{ik} - y_{jk}|^{p})^{1/p} \tag{7.11}$$

（3）曼哈顿距离（Manhattan Distance）：明氏距离中 $p=1$ 时得到的距离度量。

$$d(y_i, y_j) = \sum_{i=1}^{m} |y_{ik} - y_{jk}| \tag{7.12}$$



（4）切比雪夫距离（Chebyshev Distance）：当 p 趋向于无穷大时的明氏距离。

$$d(y_i, y_j) = \lim_{p \to \infty} \left(\sum_{k=1}^{m} |y_{ik} - y_{jk}|^p \right)^{1/p} = \max |y_i - y_j| \tag{7.13}$$

（5）马哈拉诺比斯距离（Mahalanobis Distance）：基于各指标维度进行标准化后再使用欧氏距离。

$$d(y_i, y_j) = (y_i - y_j)^{\mathrm{T}} \sum\nolimits^{-1} (y_i - y_j) \tag{7.14}$$

2）相似系数

（1）向量的夹角余弦（Cosine Similarity）：

$$r_{ij} = \frac{\sum_{k=1}^{m} y_{ik} y_{jk}}{\sqrt{\sum_{k=1}^{m} y_{ik}^2} \sqrt{\sum_{k=1}^{m} y_{jk}^2}} \tag{7.15}$$

（2）皮尔森相关系数（Pearson Correlation Coefficient）：

$$r_{ij} = \frac{\sum_{k=1}^{m} |y_{ik} - \bar{y}_i| \, |y_{jk} - \bar{a}_j|}{\sqrt{\sum_{k=1}^{m} (y_{ik} - \bar{y}_i)^2} \sqrt{\sum_{k=1}^{m} (y_{jk} - \bar{y}_j)^2}} \tag{7.16}$$

其中，$\bar{y}_i = \frac{1}{m} \sum_{k=1}^{m} y_{ik}$，$\bar{y}_j = \frac{1}{m} \sum_{k=1}^{m} y_{jk}$。

（3）Jaccard 相似系数（Jaccard Coefficient）：主要用于计算符号度量或布尔值度量的个体间的相似度，无法衡量差异具体值的大小，只关心个体间共同具有的特征是否一致这个问题。

$$\mathrm{Jaccard}(X, Y) = \frac{X \cap Y}{X \cup Y} \tag{7.17}$$

3. 传递闭包与 λ-截矩阵计算

当 X、Y、Z 为有限论域时，即 $X=\{x_1, x_2, \cdots, x_n\}$，$Y=\{y_1, y_2, \cdots, y_m\}$，$Z=\{z_1, z_2, \cdots, z_l\}$；则 $\boldsymbol{Q}$、$\boldsymbol{R}$、$\boldsymbol{S}(=\boldsymbol{Q} \circ \boldsymbol{R})$ 均可表示为矩阵形式：

$$\boldsymbol{Q} = (q_{ij})_{n \times m}, \quad \boldsymbol{R} = (r_{jk})_{m \times l}, \quad \boldsymbol{S} = (s_{ik})_{n \times l}$$

其中 $s_{ik}=\bigvee_{y\in Y}(q_{ij}\wedge r_{jk})$。$\boldsymbol{S}$ 称为模糊矩阵 $\boldsymbol{Q}$ 与 $\boldsymbol{R}$ 的乘积。

当论域为有限集时，求相似关系的等价闭包即对相似矩阵 $\boldsymbol{R}$，求 $\boldsymbol{R}^2$，$\boldsymbol{R}^4$，…。当 $\boldsymbol{R}^k\circ\boldsymbol{R}^k=\boldsymbol{R}^k$ 时，有 $e(\boldsymbol{R})=t(\boldsymbol{R})=\boldsymbol{R}^k$。

依次取 $\lambda\in[0,1]$，截关系 R_λ 是经典等价关系，它诱导出 X 上的一个划分 X/R_λ，当 λ 从 1 下降到 0 时，就得到一个划分族。由于 $\alpha>\beta$ 时，$R_\alpha[x]\subset R_\beta[x]$，即 R_α 给出的分类结果中的每类是 R_β 给出的分类结果的子类。随着 λ 的下降，R_λ 给出的分类由细到粗，形成一个动态的聚类图。

通过模糊聚类算法对信息组织进行智能聚类，主要应用在两个方面：一是在给定置信度的前提下（如 $\lambda>90\%$），通过模糊聚类过程可以智能得出最合理的分类数量；二是在给定信息组织聚类数量的前提下（如将某市的信息组织按照差异程度分成 5 类），可以同步给出该结果的可信程度（λ）。

本书附录 D 中给出了基于 MATLAB 的模糊聚类核心计算程序，包括数据标准化变换、模糊相似矩阵 $\boldsymbol{R}$ 的建立、矩阵的合成运算和动态聚类四个关键环节。

7.2.4 给定置信区间的信息协同系数及其修正形式

设定模糊聚类置信区间（对象隶属程度）的下限值 M。取 $\lambda\geqslant M$，根据模糊聚类的结果将 n 个信息组织分为 N_λ 类。

对 $k\in[1,2,\cdots,N_\lambda]$，设第 k 类组织的集合为 O_k，第 k 类组织的信息协同度为

$$z_k=\sum_{i\in O_k}\left(\sum_{j=1}^{m}y_{ij}/m\right)\tag{7.18}$$

则置信区间 $\lambda\in[M,1]$下的信息协同系数为

$$G_\lambda=1-\frac{1}{N_\lambda}\left(\sum_{i=1}^{N_\lambda-1}w_i+1\right)\tag{7.19}$$

其中 $w_i=\sum_{k=1}^{i}z_k/\sum_{k=1}^{N_\lambda}z_k$。



在信息协同系数 G_λ 的计算过程中，没有充分考虑到 n 个信息组织在置信区间 $\lambda \in [M,1]$ 下的收敛程度，因此，引入参数 $c=\frac{N_\lambda}{n}$ 对 G_λ 进行修正，修正后的信息协同系数为

$$G'_\lambda = c \times G_\lambda = \frac{N_\lambda}{n}\left(1-\frac{1}{N_\lambda}\left(\sum_{i=1}^{N_\lambda-1} w_i + 1\right)\right) \tag{7.20}$$

修正后的信息协同系数 G'_λ 将信息组织的信息协同差异性程度分为 5 个等级，对应含义如表 7.2 所示。

表 7.2　修正后的信息协同系数含义

G'_λ	<0.1	0.1～0.2	0.2～0.3	0.3～0.5	>0.5
信息协同水平	绝对平均	比较平均	相对平均	差距较大	差距悬殊

修正后的信息协同系数反映了某城市或区域内信息协同的差异水平。说明如下：

（1）区别于一般的评价过程，该评价结果给出了必要的置信度前提，且置信度与信息协同系数成反比。

（2）选择不同的距离测度方法和聚类方法会产生不同的测得结果，在实际应用中应在合理范围内综合选择多种方法进行比较，并根据实际情况对多个结果进行集结。

（3）为了保证结果的合理性，不同城市或区域间的比较应在同级别的行政区划下，在同等的聚类方法和置信区间内进行。

这里需要特别指出，关于对信息协同系数的含义解释是未来的重点研究方向之一。修正后的信息协同系数 G'_λ 将信息组织的信息协同差异性程度划分为绝对平均、比较平均、相对平均、差距较大、差距悬殊 5 个等级。对于等级划分的数量及各个级别的边界问题不在本书展开论述，感兴趣的读者可以进行深入研究，通过调研获取足够数量的国内外大中型城市样本数据，同时结合不同城市和地区的综合发展水平，建立科学的指标体系进行测算。

7.3 多源信息协同模式的横向优化策略

传统的城市系统下信息流转模式比较单一，主要是中心控制、节点参与的中心管理模式。这种模式必须由中心端发起并控制信息协同的流程，难以应对智慧城市下高并发、多数据源、高实时性的需求；同时，信息组织不能主动参与信息流转的管理过程，对信息协同状态的监控过于依赖中心端的管理，一方面随着信息协同需求的增长造成了中心端的负载过量，另一方面在信息的反应速度上严重滞后于业务需求的变化。

从信息协同网络的横向结构上，根据多源信息协同的差异测度模型，在模糊聚类的基础上对信息协同度由小到大排序的 N_λ 类信息组织进行二级层次聚类，将 n 个信息组织按照信息协同水平分成Ⅰ类组织（协同度较高）和Ⅱ类组织（协同度较低）。

设Ⅰ类组织和Ⅱ类组织分别用 O_{I} 和 O_{II} 表示，包含的信息组织数量分别为 n_{I} 和 n_{II}，需要优化的信息协同流程（包括信息发送流程和信息接收流程）数量用 δ 表示。定义实型变量 $k_{ij}\in\{0,1\}$，$i=1,2,\cdots,n$，$j=0,1,2,\cdots,n$；$k_{ij}=0$ 表示信息组织 i 与信息组织 j 不连接，$k_{ij}=1$ 表示信息组织 i 与信息组织 j 连接，其中 $j=0$ 表示信息协同中心。对信息组织 $i,j(i,j=1,2,\cdots,n)$，根据关系矩阵 $\boldsymbol{P}_A$ 和二级聚类结果进行信息协同模式优化。

对于 PCP 模式和 PCN/NCP 模式，信息组织之间不直接建立连接，$k_{ij}=0,j\neq0$。中心的信息协同负载为

$$\sum_{i=1}^{n}\sum_{j=0}^{n}(k_{i0}a_{ij}+k_{0i}a_{ji}) \tag{7.21}$$

（1）若 $i\in O_{\mathrm{I}}$，$j\in O_{\mathrm{II}}$，将 PCP 模式优化为 PCN 模式，信息组织 i 自行配置发送流程，中心配置转发流程和接收流程。设信息组织 i 的重复信息发送流程数量为 K_i，对应信息流量为 L_i，则中心的信息协同负载减少

$\sum_{i=1}^{n} K_i A_i$，可优化的信息协同流程数量为

$$\delta_{\mathrm{I}\cdot\mathrm{II}} = \sum_{i\in O_{\mathrm{I}}} \sum_{j\in O_{\mathrm{II}}} a_{ij} \tag{7.22}$$

（2）若 $i\in O_{\mathrm{II}}$，$j\in O_{\mathrm{I}}$，将 PCP 模式优化为 NCP 模式，中心配置发送流程和转发流程，信息组织 j 自行配置接收流程。设信息组织 j 的重复信息接收流程数量为 K_j，对应信息流量为 L_j，则中心的信息协同负载减少 $\sum_{j=1}^{n} K_j A_j$，可优化的信息协同流程数量为

$$\delta_{\mathrm{II}\cdot\mathrm{I}} = \sum_{i\in O_{\mathrm{II}}} \sum_{j\in O_{\mathrm{I}}} a_{ij} \tag{7.23}$$

（3）若 $i,j\in O_{\mathrm{I}}$，将 PCP 模式优化为 P2P 模式，信息组织间直接建立连接，$k_{ij}=1$，$j=0$，信息组织 i 自行配置发送流程，信息组织 j 自行配置接收流程，中心的信息协同负载减少 $\sum_{i=1}^{n}\sum_{j=1}^{n}(k_{ij}a_{ij}+k_{ji}a_{ji})$，可优化的信息协同流程数量为

$$\delta_{\mathrm{I}\cdot\mathrm{I}} = \sum_{i\in O_{\mathrm{I}}} \sum_{j\in O_{\mathrm{I}}\, j\neq i} a_{ij} \tag{7.24}$$

（4）若 $i,j\in O_{\mathrm{II}}$，保持 PCP 模式不变，由中心配置发送流程和接收流程。

信息协同网络中的信息流程总量为

$$\delta_n = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} \tag{7.25}$$

则多源信息协同结构的可优化流程占比为

$$p = (\delta_{\mathrm{I}\cdot\mathrm{I}} + 0.5\delta_{\mathrm{I}\cdot\mathrm{II}} + 0.5\delta_{\mathrm{II}\cdot\mathrm{I}})/\delta_n \tag{7.26}$$

优化后的 P2P 模式改变了原有的信息流向，减轻了中心端的负载压力，利于高频率、需求变化频繁的信息协同过程控制。优化后的 PCN 模式和 NCP 模式分别对发送端和接收端同类信息的所有流程进行了整合，即一类信息对应一个发送/接收流程，在信息内容、载体或规则发生变化时，大幅减少了协同流程的调整，提高了信息协同网络对变化的响应速度。

应用案例5　智慧城市多源信息协同的评价与优化（上）

1. 信息组织的特征提取

根据对某智慧城市基础运行领域2011—2014年的调研数据，该市城市基础运行领域主要涉及24个信息组织（政府部门、企业和社会团体），包括安全监管部门、城管执法部门、公安部门、国土部门、环保部门、交管部门、交通部门、经济信息化部门、流管部门、民防部门、气象部门、自来水集团、排水集团、电力公司、燃气集团、热力集团、市政部门、水务部门、卫生部门、消防部门、质监部门、地勘部门、应急部门、地区管理部门。其中信息提供方22个，信息需求方10个，协同信息155类。令 $n=24$，$w=155$，信息组织（对象）的集合为 $A=\{a_1,a_2,\cdots,a_{23}\}$。

信息组织的信息协同水平主要通过7个方面的特征体现：“信息需求方数量”、“共享信息种类”、“信息提供方数量”、“获取信息种类”、“信息内容种类”、“信息载体种类”、“信息周期种类”。令 $m=7$，信息组织的特征（属性）集合为 $X=\{x_1,x_2,\cdots,x_7\}$。结合信息协同的实际情况，对原始特征值进行处理，规则如下：

（1）“信息需求方数量”与“信息提供方数量”反映的均为“信息组织协同的数量”，为了突出特征间的差异性，将两类特征合并为一类，令 $x_1=x_1+x_3$。

（2）“共享信息种类”与“获取信息种类”反映的均为“信息协同的数量”，为了突出特征间的差异性，将两类特征合并为一类，令 $x_2=x_2+x_4$。

（3）“信息内容种类”主要包括基础信息、物联实时信息、视频信息、综合信息四类，令 $x_5=x_5/4$。

（4）“信息载体种类”主要包括结构化数据库、视频数据库、标准文本、普通文本、特殊文本五类，令 $x_6=x_6/5$。

（5）“信息周期种类”主要包括实时、秒、分钟、小时、天、静态六类，令 $x_7=x_7/6$。



各信息组织的原始特征值及转换后的特征值如表 7.3 所示。根据转换后的特征值构建信息协同原始矩阵 $\boldsymbol{P}$。

表 7.3　信息组织的特征值

信息组织	原始特征值							转换后的特征值				
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_1	x_2	x_3	x_4	x_5
a_1	2	41	4	34	4	4	5	6	75	1	0.8	0.83
a_2	3	2	0	0	4	1	2	3	2	1	0.2	0.33
a_3	3	6	2	7	4	3	4	5	13	1	0.6	0.67
a_4	2	1	0	0	1	1	1	2	1	0.25	0.2	0.17
a_5	6	5	1	1	1	2	2	7	6	0.25	0.4	0.33
a_6	2	1	0	0	1	1	1	2	1	0.25	0.2	0.17
a_7	4	6	6	50	3	5	5	10	56	0.75	1	0.83
a_8	2	3	4	7	3	2	3	6	10	0.75	0.4	0.5
a_9	1	1	0	0	1	1	1	1	1	0.25	0.2	0.17
a_{10}	2	1	0	0	1	1	1	2	1	0.25	0.2	0.17
a_{11}	5	33	0	0	2	2	1	5	33	0.5	0.4	0.17
a_{12}	2	3	0	0	2	1	2	2	3	0.5	0.2	0.33
a_{13}	2	2	0	0	1	1	1	2	2	0.25	0.2	0.17
a_{14}	2	2	0	0	2	1	2	2	2	0.5	0.2	0.33
a_{15}	2	2	0	0	2	1	2	2	2	0.5	0.2	0.33
a_{16}	2	2	0	0	2	1	2	2	2	0.5	0.2	0.33
a_{17}	4	5	7	15	2	2	3	11	20	0.5	0.4	0.5
a_{18}	2	12	2	3	2	1	2	4	15	0.5	0.2	0.33
a_{19}	2	5	0	0	3	1	3	2	5	0.75	0.2	0.5
a_{20}	3	3	8	49	4	5	5	11	52	1	1	0.83
a_{21}	2	8	0	0	3	2	3	2	8	0.75	0.4	0.5
a_{22}	1	11	0	0	2	1	1	1	11	0.5	0.2	0.17
a_{23}	0	0	21	136	4	5	5	21	136	1	1	0.83
a_{24}	0	0	1	4	2	2	3	1	4	0.5	0.4	0.5

2. 信息组织的一级模糊聚类

由表 7.3 可见，特征 x_1、x_2 中存在特殊值，通过公式(7.1)进行预处理。

考虑到标准差重点反映数据的离散程度，而极差重点反映数据的范围和集中趋势，因此采用平移-极差（标准 0-1）变换对数据进行标准化处理。特殊值处理和标准化处理的结果如表 7.4 所示。由表 7.4 构建信息协同标准矩阵 **Y**。

表 7.4 特殊值处理与平移-极差变换

信息组织	特殊值处理					平移-极差（标准 0-1）变换				
	x_1	x_2	x_3	x_4	x_5	x_1	x_2	x_3	x_4	x_5
a_1	0.63	0.79	1.00	0.80	0.83	0.24	0.54	1.00	0.75	1.00
a_2	0.56	0.54	1.00	0.20	0.33	0.10	0.00	1.00	0.00	0.24
a_3	0.61	0.58	1.00	0.60	0.67	0.20	0.09	1.00	0.50	0.76
a_4	0.53	0.54	0.25	0.20	0.17	0.04	0.00	0.00	0.00	0.00
a_5	0.66	0.55	0.25	0.40	0.33	0.31	0.02	0.00	0.25	0.24
a_6	0.53	0.54	0.25	0.20	0.17	0.04	0.00	0.00	0.00	0.00
a_7	0.73	0.73	0.75	1.00	0.83	0.45	0.41	0.67	1.00	1.00
a_8	0.63	0.57	0.75	0.40	0.50	0.24	0.07	0.67	0.25	0.50
a_9	0.51	0.54	0.25	0.20	0.17	0.00	0.00	0.00	0.00	0.00
a_{10}	0.53	0.54	0.25	0.20	0.17	0.04	0.00	0.00	0.00	0.00
a_{11}	0.61	0.65	0.50	0.40	0.17	0.20	0.24	0.33	0.25	0.00
a_{12}	0.53	0.54	0.50	0.20	0.33	0.04	0.00	0.33	0.00	0.24
a_{13}	0.53	0.54	0.25	0.20	0.17	0.04	0.00	0.00	0.00	0.00
a_{14}	0.53	0.54	0.50	0.20	0.33	0.04	0.00	0.33	0.00	0.24
a_{15}	0.53	0.54	0.50	0.20	0.33	0.04	0.00	0.33	0.00	0.24
a_{16}	0.53	0.54	0.50	0.20	0.33	0.04	0.00	0.33	0.00	0.24
a_{17}	0.75	0.60	0.50	0.40	0.50	0.49	0.13	0.33	0.25	0.50
a_{18}	0.58	0.59	0.50	0.20	0.33	0.14	0.11	0.33	0.00	0.24
a_{19}	0.53	0.55	0.75	0.20	0.50	0.04	0.02	0.67	0.00	0.50
a_{20}	0.75	0.71	1.00	1.00	0.83	0.49	0.37	1.00	1.00	1.00
a_{21}	0.53	0.56	0.75	0.40	0.50	0.04	0.04	0.67	0.25	0.50
a_{22}	0.51	0.57	0.50	0.20	0.17	0.00	0.07	0.33	0.00	0.00
a_{23}	1.00	1.00	1.00	1.00	0.83	1.00	1.00	1.00	1.00	1.00
a_{24}	0.51	0.55	0.50	0.40	0.50	0.00	0.02	0.33	0.25	0.50

由于各特征均为效益型变量，选择距离测度中的欧式距离和切比雪夫距离进行构建模糊相似矩阵 $\boldsymbol{R}$（详细数据见附录 E 的表 E.1 和表 E.2）。计算传递闭包 B（详细数据见附录 E 的表 E.3 和表 E.4）和 λ-截矩阵 $\boldsymbol{D}$，分别经过 18 次（基于欧氏距离）和 11 次（基于切比雪夫距离）迭代，得到模糊聚类的动态过程，如图 7.1 所示。

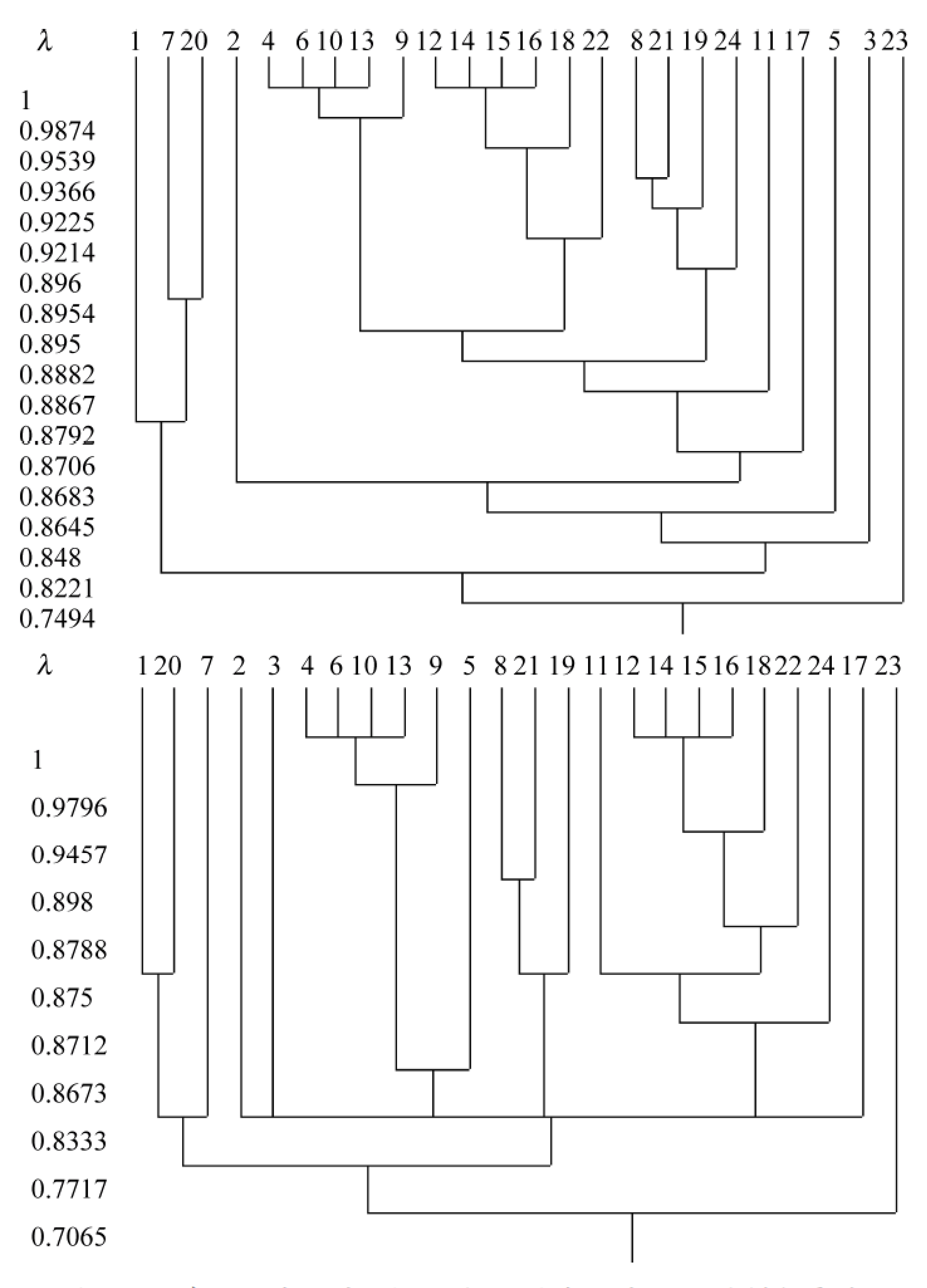


图 7.1　基于欧式距离（上）和切比雪夫距离（下）的模糊聚类

根据图 7.1 可以选定置信度，自动将该市城市基础运行领域的信息组织分成若干类；或选定信息组织聚类的数量，同步给出该分类结果的置信程

度。在此，以欧式距离为测度标准，取置信度高于89%的模糊聚类结果，即$\lambda \geqslant 0.89$，根据模糊聚类的结果将24个信息组织分为10类，即$N_{0.89}=10$。按类的信息协同度由小到大排序，如表7.5所示。

表7.5　置信度>0.89的模糊聚类结果

聚　类	包含的信息组织	聚　类	包含的信息组织
第1类	$a_4, a_6, a_9, a_{10}, a_{12}, a_{13}, a_{14}, a_{15}, a_{16}, a_{18}, a_{22}$	第6类	a_{17}
第2类	a_5	第7类	a_3
第3类	a_{11}	第8类	a_1
第4类	a_2	第9类	a_7, a_{20}
第5类	$a_8, a_{19}, a_{21}, a_{24}$	第10类	a_{23}

3. 给定置信区间的信息协同差异测度

令$z_k(k=1,2,\cdots,10)$表示第k类组织的信息协同度，结果如表7.6所示。

表7.6　置信度>0.89的类协同度

z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	z_9	z_{10}
0.07	0.16	0.21	0.27	0.28	0.34	0.51	0.71	0.74	1.00

对应w_i的结果如表7.7所示。

表7.7　置信度>0.89的类权重系数

w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}
0.02	0.05	0.10	0.17	0.23	0.31	0.43	0.60	0.77	1.00

则置信区间$\lambda \in [0.89,1]$下的信息协同系数为

$$G_{0.89}=1-\frac{1}{10}\left(\sum_{i=1}^{9} w_i+1\right)=0.6323$$

引入修正参数$c=\frac{10}{24}\approx 0.417$对$G_\lambda$进行修正，修正后的信息协同系数为



$$G'_{0.89} = c \times G_{0.89} = 0.263$$

根据表 7.2，$G'_{0.89} \in [0.2, 0.3]$说明在 0.89 的置信度下该市城市基础运行领域的信息协同水平相对平均。信息协同曲线如图 7.2 所示。

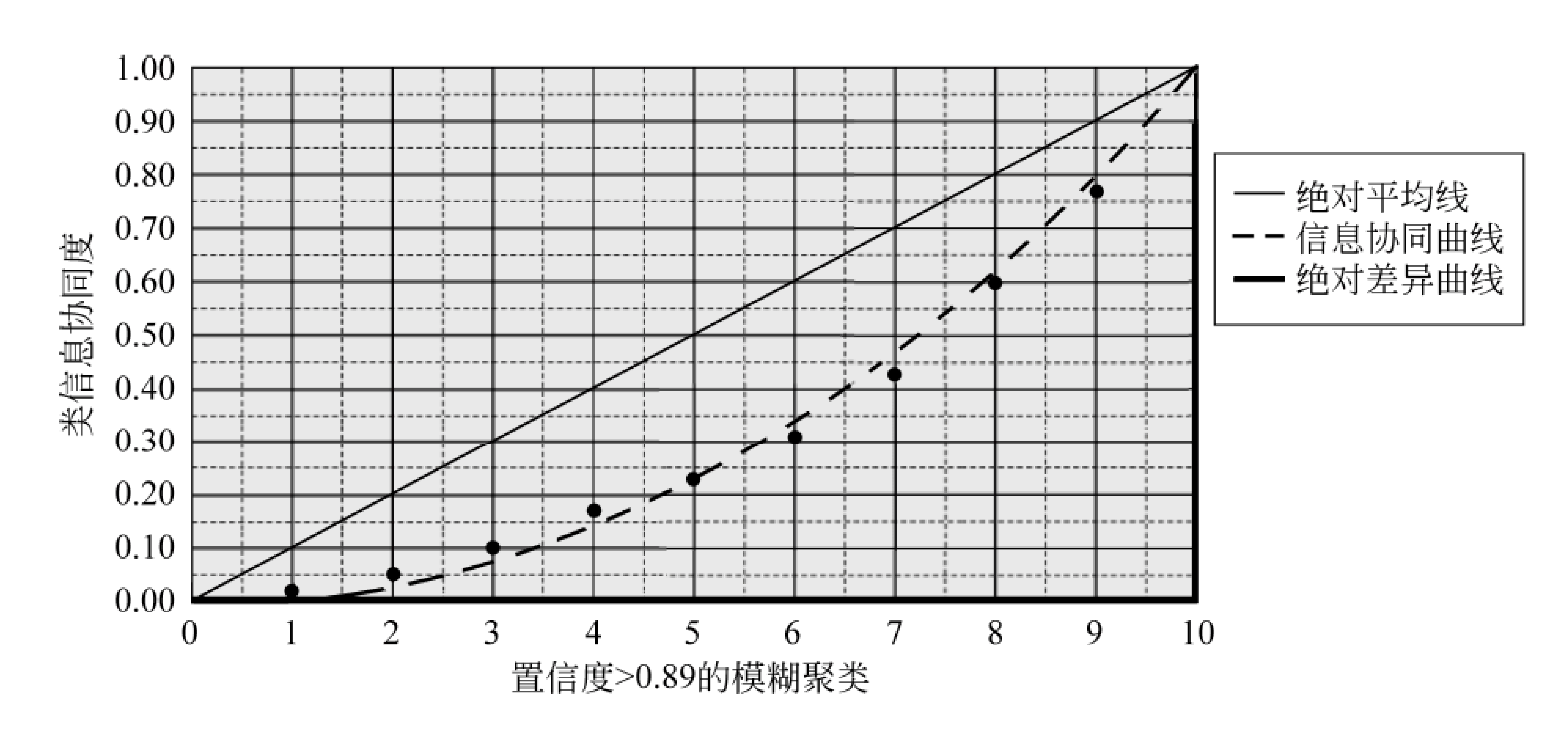


图 7.2　某智慧城市基础运行领域的信息协同曲线

在图 7.2 中，信息协同曲线与绝对平均线之间的面积表示城市信息协同水平的差异程度，即修正后的信息协同系数的几何意义。面积越小（即信息协同曲线弧度越小），说明信息协同水平的差异越大，反之则差异越小。

4. 信息组织的二级聚类与信息流转模式优化

通过层次聚类方法，对信息协同度由小到大排序的 10 类信息组织进行二级聚类，结果如图 7.3 所示。

由图 7.3 可见，该市城市基础运行领域的 24 个信息组织按照信息协同水平可以分为两大类，如表 7.8 所示。

表 7.8　信息组织分类

一级聚类	二级聚类	包含的信息组织
第 1～6 类	Ⅱ类	a_2，a_4，a_5，a_6，a_8，a_9，a_{10}，a_{11}，a_{12}，a_{13}，a_{14}，a_{15}，a_{16}，a_{17}，a_{18}，a_{19}，a_{21}，a_{22}，a_{24}
第 7～10 类	Ⅰ类	a_1，a_3，a_7，a_{20}，a_{23}

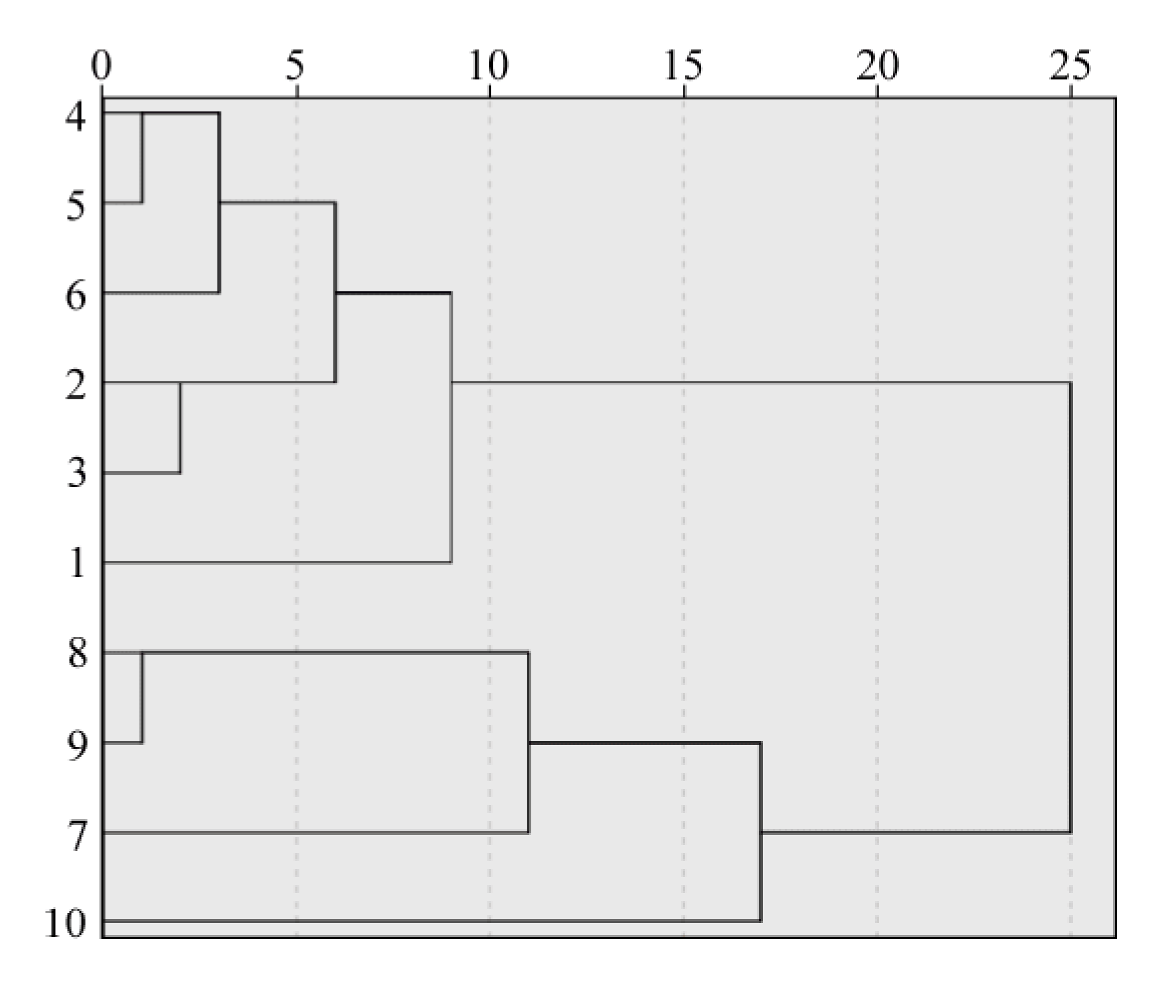


图 7.3 信息组织的二级层次聚类

根据二级聚类结果和模式优化策略，组织 a_1，a_3，a_7，a_{20}，a_{23} 的信息协同水平较高，相互之间的信息协同采用 P2P 模式；对组织 a_2，a_4，a_5，a_6，a_8，a_9，a_{10}，a_{11}，a_{12}，a_{13}，a_{14}，a_{15}，a_{16}，a_{17}，a_{18}，a_{19}，a_{21}，a_{22}，a_{24} 之间的信息协同仍然采用传统的 PCP 模式，对分属于Ⅰ类和Ⅱ类的组织间的信息协同采用 PCN 或 NCP 模式。

根据公式(7.21)～公式(7.26)计算可得：

$$\delta_n = 306, \quad \delta_{\mathrm{I} \cdot \mathrm{I}} = 69, \quad \delta_{\mathrm{I} \cdot \mathrm{II}} = 8, \quad \delta_{\mathrm{II} \cdot \mathrm{I}} = 205$$

信息协同结构的可优化流程占比 p 为 57.35%，即对现有的半数以上信息协同流程实现了优化。同属Ⅰ类组织的 69 类信息流转不再通过中心端，而是直接建立对接关系；发送方属于Ⅰ类组织、接收方属于Ⅱ类组织的 8 类信息将重复发送流程约减为统一的 8 个流程，通过中心端向不同的接收方分发；接收方属于Ⅰ类组织、发送方属于Ⅱ类组织的 205 类信息将重复接收流程约减为统一的 205 个流程，通过中心端自行接收。

优化后的 P2P、PCN 和 NCP 模式改变了原有 PCP 模式下单一的信息流向，减轻了中心端的负载压力；同时，当外界需求发生变化时，减少了对流程的人工干预，提高了信息协同网络对变化的自动调整能力和响应速度。

第 8 章　基于凝聚子群的协同网络关系测度

差异测度主要利用的是信息组织的“属性变量”，体现协同结构中信息组织的特征相似程度。对协同结构的关系测度需要利用信息组织间的“关系变量”，分析协同网络中存在的凝聚子群及其相互关系。

凝聚子群的研究属于社会网络分析（Social Network Analysis，SNA）范畴。社会网络分析是新经济社会学中一种重要的研究范式，它认为互动的成员间存在的关系非常重要，力图用图论工具、代数模型技术描述关系模式，并探究这些关系模式对结构中的成员或整体的影响。社会网络分析的核心在于从“关系”的角度出发研究社会现象和社会结构，其中社会结构可以是行为结构、政治结构、经济结构等多种形式。



关于社会网络分析的基本理论和 UCINET 软件的应用，国内的刘军教授已经做了非常专业和详细的论述，在此不再赘述。考虑到本书理论体系的完整性，只对相关内容进行总体性介绍，并给出一个具体的应用案例供读者参考。

8.1　社会网络分析

8.1.1　社会网络分析的基础理论[①]

1. 社会网络分析的基本概念和假设

网络是由事物以及事物之间的某种关系构成的，对事物结构的关注可以看成是一种网络视角。社会网络指的是社会行动者及其间的关系的集合。

① 感兴趣的读者可进一步参阅刘军教授的《社会网络分析导论》和《整体网分析讲义——UCINET 软件实用指南》，详细信息请在本书参考文献中查阅。

一个社会网络是由多个点（社会行动者）和各点之间的连线（行动者之间的关系）组成的集合。用点和线来表达网络，是社会网络的形式化界定。

1）点（社会行动者）

社会网络分析中的行动者可以是任何一个社会实体（单位或个人），如学校、村落、组织、城市、国家等。点可以是任何社会行动者，关于点的信息必须是实际信息，可用常规方法进行收集。信息可以是动态的，也可以是静态的。

2）关系（行动者之间的联系）

一般来说，当我们说行动者之间存在关系（ties）的时候，"关系"常常代表的是关系的具体内容（relational content）或者是实质性的现实发生的关系。关系有多种表现：

首先，行动者之间的关系类型多样，如朋友关系、上下级关系、国家之间的贸易关系、城市之间的距离关系等。

其次，行动者之间存在"多元关系"，如两个国家之间可能存在贸易关系、外交关系、文化往来关系等。对多元关系网络的研究，特别是整体网模型研究是当今社会网络分析中最具潜力的前沿领域。社会网络研究者利用多维量表（MDS）、矩阵代数（Matrix Algebra）、聚类分析（Cluster Analysis）等多种方法来研究多元关系网络数据。也有很多学者利用概率论、数理统计技术以及计算机技术研究网络变量的统计性质，构建多种网络模型。

再次，研究的重点不同，关注的"关系"也不同。如果研究整体网络（whole network），即研究所有行动者之间的关系，那么研究者需要分析具有整体意义的关系的各种特征，如互惠性、关系的传递性等。如果研究个体网络（ego-network），即关注个体行动者，则需要分析个体网的一些关系特征，例如，关系的密度、同质性等。这种研究可以利用随机抽样方法。

2. 社会网络分析的假设前提

社会网络分析建立在如下假设基础之上：在互动的单位之间存在的



关系非常重要，关系是网络分析理论的基础。网络学者坚持如下前提性论题：

（1）行动者以及行动是相互依赖的，而不是独立的、自主性的单位。

（2）关注行动者之间的关系（而不是属性），行动者之间的关系是资源（物质的或者非物质的）传递或者流动的“渠道”。

（3）个体网络模型认为，网络结构环境可以为个体的行动提供机会，也可能限制其行动，多元行动者之间的关系会影响到人们的行为。

（4）网络模型把结构（社会结构、经济结构等）概念化为各个行动者之间的关系模型，“结构是网络之网”。



3. 社会网络分析的研究视角

在认识论上，社会网络分析认为世界是由网络而不是由群体组成的。从网络而不是群体出发，可以把世界看成是网络的结构，把行动者之间的关系看成是资源流动的渠道，从而可以通过分析发现复杂的资源流动网络，而不是简单的分层结构。这样，我们就应该根据行动者之间的关系模式来理解观察到的社会行动者的属性特征（如种族、生产总量等）。行动者之间的关系居于首要地位，而行动者的属性居于次要地位。

在方法论上，社会网络分析认为从关系视角进行的解释要优越于从属性视角进行的解释，“网络理论把解释建立在关系模式之上”（Emirbayer，1994）。

许多主流社会学研究把社会结构和过程看成是个体行动者的个人属性的总和。不管这些属性是先赋的（性别、智力等）还是自致的（社会经济地位、政治态度等），都被看成是个体的属性因素。每种属性都看成是独立的分析单位，对这些因素进行分析的各种统计方法（各种多变量分析技术）都把拥有相同属性的个体归为同一类，从而作为分析的单位。

这种分类分析把个体归为具有相同属性的类，因此没有考虑到个体所处的关系结构。此类研究认为，社会行为之所以出现，是因为个体拥有与其

他个体相似的属性，而不是因为个体处于一定的社会关系网络之中。尽管很多社会学家声称要通过结构研究考察社会关系，但是，他们所使用的结构技巧以及研究方法论关注的仍然是变量之间的结构，而这种结构很可能只是研究者自己建构出来的，不是行动者之间的真实结构。这种分析有如下问题：一是分析的重点是行动者的属性特征；二是属性分析把每个社会系统成员看成是非结构性的相互独立的单位；三是把社会结构解释为受规范引导的现象，这种分析破坏了结构关联的信息；四是当考察类别的时候，一般认为同类人的行为方式也相同。

网络结构分析可以为分析社会结构提供精致的工具。网络分析者认为，行动者既属于一定的类别，又处于一定的网络之中，因而不能仅仅考察其属性，还应该关注其所处的关系网络。例如，不能仅仅把社会阶级看成是一个地位集合，还应该看成是权力的经济关系和依赖性的综合，所以，从社会行动者所处的社会关系的角度进行的解释和研究是比较优越的。当然，网络研究者并不排斥“属性”研究的正当性。

社会网络分析的解释不同于非网络解释之处在于：在研究中把关于“关系”的概念和过程纳入解释之中。社会网络研究涉及的理论概念是关系性的，相关数据是关系性的，对数据的假设检验也使用关系属性的分布。无论利用的模型是为了理解关系背景下的个体行动，还是直接研究结构，网络分析都根据“关系”对结构进行操作。互动的模式就是结构，而“标准”的社会科学量化视角常常忽视关系性的互动结构。

8.1.2 社会网络分析的数据类型和研究方法

1. 社会网络分析的数据类型

社会网络分析的数据主要分为属性数据、关系数据和观念数据三类。

（1）属性数据：关于行动者的自然情况、态度、观点以及行为等方面的数据，一般被视为个人或者群体所具有的财产、性质、特点等属性，是人们、

对象或者事件的内在特点，适用于分析属性数据的方法主要是变量分析，如相关分析、回归分析、列联表分析等。各种属性被看成是特定变量（收入、职业、教育程度等）的取值。

（2）关系数据：关于联系、接触、联络等方面的数据。这类数据把一个行动者与另一个行动者连接在一起，因此不能还原为单个行动者的属性。关系不是行动者的属性，而是行动者系统的属性，这些关系把多对行动者联系成一个更大的关系系统。

传统数据关注的是行动者和属性，网络数据关注的是行动者和关系。相比常规的定量统计分析，社会网络分析方法更适用于分析关系数据。关系数据主要包括两种类型：一是行动者-行动者数据（方阵数据），即 1-模网络数据；二是行动者-事件（长方阵数据），即隶属关系数据。

（3）观念数据：主要描述意义、动机、定义等，目前比较有代表性的研究方法是类型分析。



2. 社会网络分析的研究重点和方法

1）密度

密度是社会网络分析中最常用的一种测度，它在社会网络分析中占据重要地位。具体来说，密度指的是网络中各个成员之间联系的紧密程度，具体数值是通过网络中实际存在的关系数量与理论上可能存在的关系数量（完备图）相比得到。成员之间的联系越多，该网络的密度也就越大。

2）网络中心度和中心势

“中心性”是社会网络分析中的重点，成员在其社会网络中具有怎样的权力是社会网络分析者最早探讨的内容之一，它们从关系的角度定量地界定权力，并通过中心度和中心势指数进行测量。中心度是衡量成员处于网络中心位置的程度，其中点度中心度最常用，通过计算它与多少点直接相连得到，数值大表明该成员处于中心地位。中心势是度量整个网络中心化的

程度，测量网络的总体整合度或者一致性，如星形网络，所有成员只围绕一个成员发生联系，其他成员间都没有联系，这样网络的中心势最高。

3）凝聚子群

社会结构是在社会成员之间实存或者潜在的关系模式。它主要研究网络中存在多少凝聚子群，各个凝聚子群间是什么关系，以及凝聚子群内部成员之间的关系具有怎样的特点等。目前，在社会网络研究中，还没有比较明确的凝聚子群的定义，大体就是指成员之间具有相对较强的、直接的、紧密的、经常的或者积极的关系所构成的一个成员的子集合。如果网络中存在较多的凝聚子群，并且这些凝聚子群间缺少交往，则这样的关系结构不利于整体网络的发展。

4）结构相似性

结构相似性分析能够对社会行为和社会结构进行一般化分析，更好地把握成员之间关系模式的相似性，得到具有推广意义的结论。在一种网络关系中，如果两个成员相互替代后不改变整个网络的结构，就说明这两个成员具有结构相似性。一般除了与自身外，完全的结构相似是很少见的。结构相似性和凝聚子群的分析是不同的，前者的目的是把相似的成员分到互斥的群体中，每一群体内部的成员之间拥有类似的结构特征，它们是相互对等的，但各个群体中的结点间是不对等的；凝聚子群的研究是为了找到整体网络中的一些子群体。

5）核心-边缘结构

“核心-边缘”结构的分析范式已经成为区域经济学中一种相对固定的模式，随着区域经济的发展，城市间的关系变得日益复杂，中心区、半边缘以及边缘地区不断变更，这种变动会影响区域经济的发展。因此，有必要清楚网络结构中是否存在核心——边缘结构，以及中心城市在城市群中的经济位置是什么。

应用案例6 基于改进重力模型的省际流动人口的复杂网络分析[①]



1. 背景介绍

流动人口是中国户籍制度的产物，指以工作、生活为目的，离开户籍所在地，到其他地方居住的人口，其中省际人口流动是指跨越省界的长距离流动。改革开放以来，我国人口出现大规模、多层次的流动，人口流动呈现逐年递增状态，人口流动量不断增大，流动形式多样化。人口流动是地区间、城乡间和产业间市场经济资源优化配置的内在要求，是实现人口现代化和劳动价值规律的必然现象。省际之间的人口流动对于活化区域间社会经济系统，改善地区间经济发展不平衡和缩减收益差距具有重要作用。因此对人口的流动规律研究和管理一直是学术界和政府部门关注的热点，逐渐引起各级政府机构以及管理部门的重视。

国内外学者对流动人口进行了大量的研究。在研究视角上，大部分的研究主要集中于人口迁移的因素分析、空间分布特征、政策制定与研究等方面，而且多聚焦城乡流动人口分布。励娜等采用多元回归分析的方法，分析中国城乡人口流动的时间和区域趋势及驱动因素。乔晓春等利用2010年第六次全国人口普查数据，对不同省市自治区的跨省流动人口和分省户籍人口分布状况，以及分省迁入率和迁出率进行了估计。宋健等对1984年以来北京市流动人口管理的相关政策法规进行了整理、分类和分析，从户口管理、住房管理、计划生育管理和就业管理等方面作了具体阐述。在研究方法上针对流动人口研究较为早期的主要集中于描述分析、数理统计方法以及组合数学模型等。其中侯贺平等采用改进辐射模型，从节点和社团结构以及无标度和小世界特征等方面，探讨在地域差异影响下人口流动的空间分

① 本案例由中国科学院王宁宁博士提供，详细信息可查阅本书参考文献。

布格局和复杂网络特征。鲍常勇借助描述统计方法，对我国286个地级及以上城市流动人口分布特征进行了分析。随着复杂科学和复杂网络的研究兴起，复杂网络分析方法被应用到人口流动研究领域，为研究提供了崭新的视角和方法。

复杂网络起源于匈牙利数学家Erdos和Renyi的随机图理论，随着无标度网络和小世界网络的出现，复杂网络的研究开始进入新的阶段。复杂网络主要研究的是个体之间相互作用所产生的系统的整体性质与行为，从而揭示和把握复杂系统的宏观特征。社会网络分析是复杂网络分析的方法之一，是从网络的视角出发，探求社会行动者以及他们之间关系的研究方法，它起源于物理学中的适应性网络，通过研究网络关系，把个体间关系、微观网络与大规模的社会系统宏观结构结合起来，采用图论等定量的数学分析方法，解决社会问题，该方法在职业流动、城市问题、人口流动以及世界政治经济体系等领域广泛应用。

已有的人口复杂网络研究多以实际的人口流动建立网络，聚焦人口流动网络的拓扑结构，从而研究人口流动的空间分布特征和流动趋势。针对潜在未来可能发生的人口流动预期网络研究还处于相对薄弱的状态，多数的预期的流动人口研究包括可达性的研究以及经典的人口重力模型与空间相互作用机会模型研究。鉴于此，本案例首先在整体上根据省际真实的人口流动情况建立网络，形成整体的流动人口网络视图，并且探讨了真实流动人口网络的拓扑结构和网络的相关指标和特征，然后对比分析了基于交通成本的流动人口测度网络和传统重力模型的流动人口测度网络，在此基础上，对经典的重力模型进行修正，建立了基于经济、就业以及交通的人口流动重力测度模型，借助社会网络分析（SNA）方法，依托UCINET社会网络分析软件，深入研究省际预期潜在流动人口网络，分析流动人口网络的整体特征，找出网络中的核心省市、核心区域、纽带省市以及人口流动省市子群，根据网络的特征，促进省际人口资源的有效流通和共享，为流动人口管理以及产业布局提供科学决策。



2.省际流动人口的网络拓扑结构分析

本案例涉及的人口流动是从户籍角度出发，将户籍地在外省市的人口定义为流动人口。根据第六次全国人口普查数据，全国 31 个省以及直辖市之间都存在人口流动，因此在此基础上的流动人口网络是完备网络，即 31 个网络节点之间都存在联系。为了更加清楚直观地了解省际流动人口网络的拓扑特征，对流动人口量设定相关阈值，当省际流动人口数量在 10 000 以上时定义为省际之间存在大规模流动人口，此时节点省或直辖市之间存在人口流动联系。

根据第六次全国人口普查数据建立人口流动矩阵，采用社会网络 UCINET 分析，得到出度与入度最大的前十个省（自治区）（见表 8.1）。从表 8.1 和图 8.1 可以看出，人口流出省（自治区）主要集中在中部经济水平较为不发达的人口大省，其中安徽和河南是最典型的人口流出省，与劳动力输出有直接的关系。人口流入大省主要集中在北上广以及东部沿海等经济较为发达的省份，值得关注的是新疆在流入人口中的数量排名也比较靠前，这与国家援助新疆建设的人口政策相关。

表 8.1　流动人口排名前十的省（自治区）

排名	省（自治区）（流出人口）	省（自治区）（流入人口）	排名	省（自治区）（流出人口）	省（自治区）（流入人口）
1	安徽	广东	6	江西	福建
2	河南	上海	7	广西	天津
3	四川	北京	8	河北	山东
4	湖南	浙江	9	江苏	辽宁
5	湖北	江苏	10	山东	新疆

（数据来源：第六次全国人口普查）

度反映的是网络的中心性程度，特征途径路径长度是网络整体性质的测度指标。大规模省际流动人口网络是有向网络，具有 31 个省市节点，共有 1024 条边，具体的网络特征统计如表 8.2 所示，每个网络节点指向其他节点

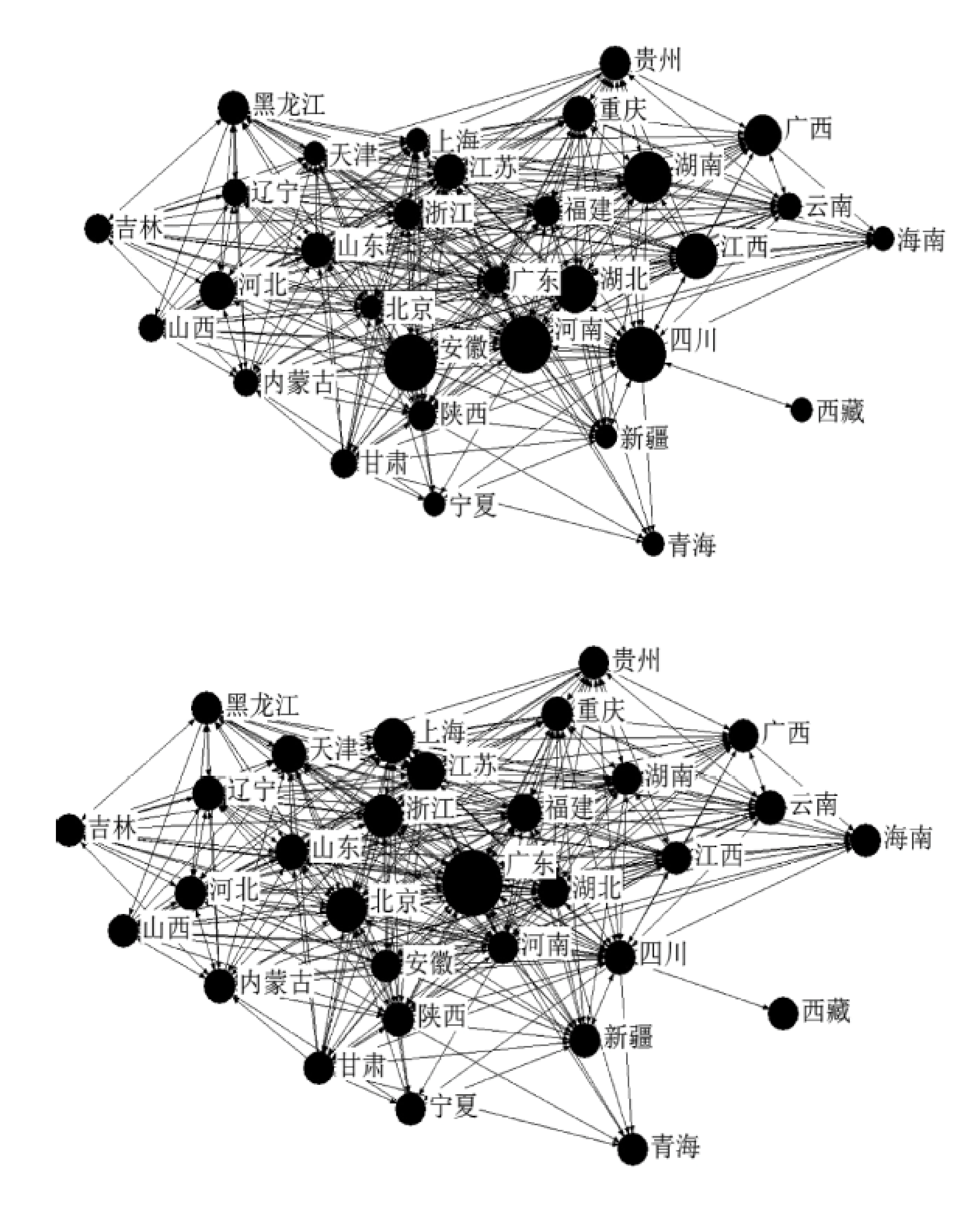


图 8.1 省际流动人口网络属性图(左为出度，右为入度，

圆点大小表示中心度相对大小)

的边的数量在 1～29 之间，而指向每个节点的边的数量在 1～28 之间，略有差异，但是平均值都为 14.935，说明大规模流动人口网络的中心性较为明显，存在度较大的网络权利节点。网络的平均距离为 1.571，网络中任意两点之间的平均距离位于 1～2 之间，大部分的省节点通过 1～2 个中间省建立流动人口联系，平均经过 1.571 距离发生人口流动，大规模省际流动人口网络呈现小世界的特征。

表 8.2 省际流动人口网络相关指标值

节点	边	最大出度	最小出度	平均出度	最大入度	最小入度	平均入度	平均距离
31	1024	29	1	14.935	28	1	14.935	1.571

度分布反映的是网络节点性质的宏观统计特征。设定阈值的大规模流



动人口网络是有向网络，具有 31 个省市节点，共有 1024 条边。由于网络的规模相对较小，因此对网络进行对称化处理，采用累积概率分布描述大规模流动人口网络的宏观统计特征，其具体的计算公式如下所示：

$$P(k) = \sum_{k'=k}^{\infty} p(k)$$

图 8.2 描述了设定阈值的大规模流动人口网络的累计概率分布(双对数处理)，对 31 个节点度以及概率进行幂函数拟合得到幂分布函数：$P(k) = 2.208\,933k^{-0.544\,26}$，说明网络的度累计概率密度与度数呈现一定的幂律关系，但是指数 λ 为 $0.544\,26<2$，大规模流动人口网络的无标度特征不明显，度分布指数较小，流动人口网络的异质性比较高，网络中不同度值的节点分布均匀。

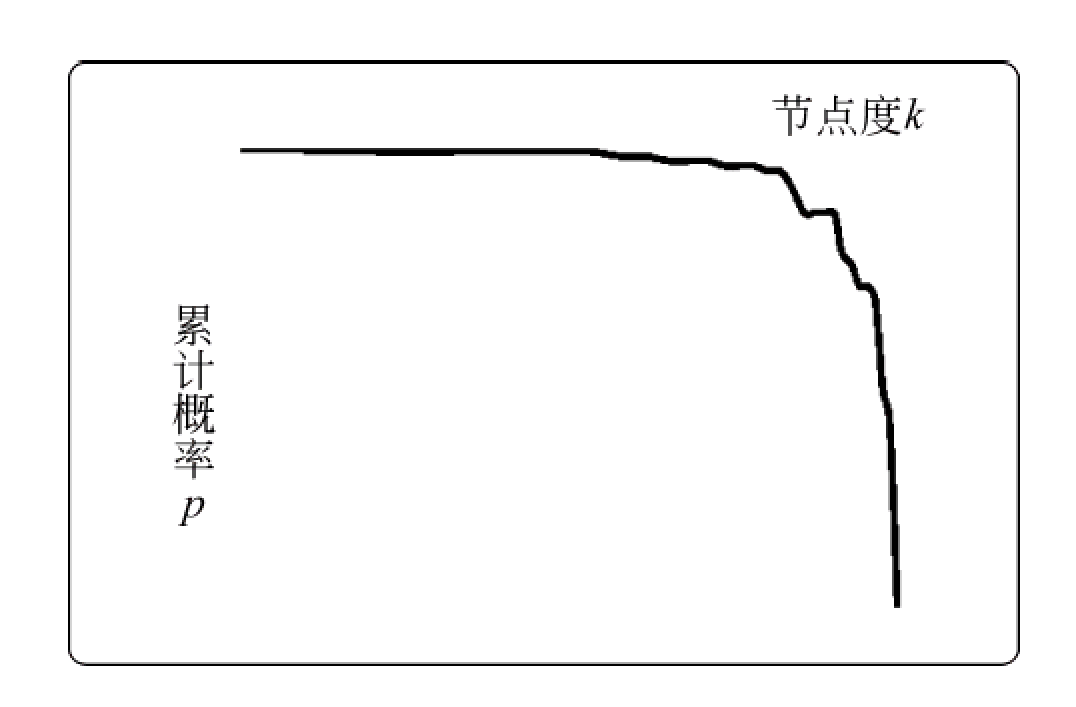


图 8.2　大规模流动人口网络的累计概率分布(双对数)

3. 基于流动人口数量的权重网络分析

省际流动人口网络可以看作以流动人口数量为权重的权重网络，权重 w_{ij} 为节点省市 i 到 j 的流动人口数量。由于省际流动人口网络是有向网络，所以通常情况下 $w_{ij} \neq w_{ji}$。为了更好地研究流动人口网络的特征，参考第 2 部分设置阈值 10 000 建立大规模的权重流动人口网络模型。

1) 权重网络的度分析

权重网络的节点度计算参考公式 $D = \sum_{j \in v(i)} w_{ij}$，其中 $v(i)$ 是节点 i 的邻居节点集合。根据人口流动网络的实际意义，省际流动人口权重网络的入度

和出度可以分别定义为人口流入强度 S_{in} 和人口流出强度 S_{out}。

表 8.3 分别列出了人口流入强度和人口流出强度的前十个省，排名显示人口流入最强的省市主要集中分布于中国的东部沿海地区以及京津冀等地区，这与该地区的经济发展水平较高密切相关；此外新疆地区的人口流入强度也相对较为突出，与国家鼓励西部大开发以及援疆建设政策息息相关。人口流出强度大的地区主要分布于中国的中部地区，包括安徽、河南、湖南、湖北等省。究其原因，一方面经济发展相对薄弱，更多的人口为追求更好的生活而外出打工；另一方面人口数量庞大，劳动力富裕。

表 8.3　人口流动权重网络的度分析

排名	人口流入强度	人口流出强度	排名	人口流入强度	人口流出强度
1	广东	安徽	6	福建	江西
2	上海	河南	7	天津	广西
3	北京	四川	8	山东	河北
4	浙江	湖南	9	辽宁	江苏
5	江苏	湖北	10	新疆	山东

2）权重网络的度与经济人口的相关性分析

为了进一步验证省市人口流动强度与地区经济和人口总量的相关关系，以地区人均 GDP 和人口总量为自变量，分别对人口流出强度和人口流入强度进行简单线性回归（为了减少多重共线性和消除量纲的影响，对原始数据分布作对数处理），回归结果见表 8.4。

表 8.4　人口流动强度回归统计指标及结果

	回 归 方 程	R	F	Significance F	t-人均 GDP	t-人口总量
流出人口强度回归	$S_{out}=-0.97GDP(pc)+1.82Po+9.12$	0.90387	62.4976	4.72851E-11	−2.855159	11.00141
流出人口强度回归	$S_{in}=2.495GDP(pc)+0.625Po-18.09$	0.87452	45.5204	1.5866E-09	8.1880054	4.235571

人口流动强度的回归方程中，R 约为 0.9，回归方程的拟合度较好，回归方程的 F 统计量均远远大于对应的 P 值，说明人口流动强度（流出强度和流入强度）与地区人均 GDP 和人口总量具有显著的相关性。相关系数检验显著，人口流出强度与地区人均 GDP 呈现反向线性相关，与地区人口总量正向线性相关，较低的地区人均 GDP 和较大人口总量是人口流出地区的推力。人口流入强度与地区人均 GDP 和人口总量正向线性相关，说明一个地区的较高经济发展水平和较大人口总量具有对外来人口较强的吸引力。

4. 流动人口的测度模型与复杂网络构建

1）基于交通成本的测度

（1）测度模型描述

人口在流动的时候通常需要考虑的是交通成本，包括交通时间和旅行费用等。通常情况下，在人口流入地的经济吸引力相似的情况下，交通成本越低的地区越容易吸引大量的人口流入。

由于交通时间和旅行费用与距离以及交通工具的选择有关，因此人口总是倾向于向距离较近的省（市）迁移。定义 S_{ij}^{T} 为省 i 与省 j 之间基于交通成本的预期人口流动强度。具体的计算公式如下：

$$S_{ij}^{\mathrm{T}} = \frac{1}{d_{ij}}$$

其中，d_{ij} 为省 i 与省 j 的空间距离，考虑到各省之间距离统计的难度，以省会城市之间的铁路里程替代，根据 S_{ij}^{T} 的计算公式，可以进一步计算基于交通成本的省市人口流动可达性 $A_i^{\mathrm{T}} = \sum\limits_{j \neq i} \frac{1}{d_{ij}}$。$A_i^{\mathrm{T}}$ 表示在考虑距离层面上省 i 对其他省市人口的引力强度。

基于交通成本的测度模型主要是从可达性的角度，以各省之间的距离为主要的测量指标来衡量人口流动的强度，指标计算简单，由于距离不存在方向性，因此建立的人口流动强度是无向的，只能粗略反映省际人口流动情况。

（2）基于交通成本的人口流动可达性与流动网络

为了简化计算，本案例将各省看成以省会为中心的理想的质点。以省会城市之间的铁路距离为省之间的距离，计算各省的人口流动可达性。表8.5列出了计算结果中可达性排名前十的省（市）以及实际流入人口数量排名前十的省（市），图8.3直观显示了可达性结果与实际的对比情况。

表8.5 可达性计算结果与实际流入人口排名

排名	省（市）	流入人口	省（市）	可达性
1	广东	17632894	安徽	0.035174
2	上海	6361198	河北	0.033228
3	北京	5869769	北京	0.032995
4	浙江	5728950	江苏	0.032834
5	江苏	4208538	天津	0.03179
6	福建	2708530	湖北	0.031101
7	天津	2459656	山东	0.030208
8	山东	1542925	河南	0.02924
9	辽宁	1405459	上海	0.028784

基于交通成本计算出的流动人口可达性前十名的省（市）中只有五省（市）进入到实际发生的流动人口省（市）前十名，分别是北京、上海、江苏、天津和山东，比例占到50%。这个比例相对来说比较低。根据图8.3也可以明显地看出，实际的人口流动大省与基于交通成本计算出来的人口流动大省差异明显，实际的主要集中在中东部，而计算得出的省份主要分布在东南沿海以及东北地区。这说明交通成本虽然是影响省际人口流动的因素，但不是关键、唯一的因素。从以往的研究以及文献中也可以得出，影响省际人口流动的因素除了交通成本（距离和费用）外，经济发展水平以及地区的人口数量、社会文化因素等都是人口流动的重要因素。

2）基于重力模型的测度

（1）测度模型描述

重力模型起源于物理学领域的引力模型，作为描述人口迁移的经典简



单数学模型，其出发点是假设迁移流取决于出发地与接收地的人口数及两地之间的地理距离，大多数迁移所越过的距离很短，当吸收迁移的中心地点距离迁出地愈远时，迁移人数迅速下降。基于重力模型的人口流动测度公式如下：

$$S_{ij}^{G} = k\frac{Q_i Q_j}{d_{ij}^{b}}$$

其中，S_{ij}^{G}表示省 i 与省 j 之间的人口引力大小，k 为重力系数，Q_i 与 Q_j 分布表示省 i 与省 j 的人口总量，d_{ij} 为省 i 与省 j 之间的交通距离，b 是距离衰减系数，衰减系数一般在 1～2 之间。

（2）基于重力模型的人口流动网络

根据上述重力模型公式计算省际之间的人口引力，构建引力矩阵。省际之间的距离采用省会城市之间的距离，城市之间的距离统计是按照中国铁路通达性进行计算的，主要基于以下考虑，中国的铁路覆盖面积比较广，城市之间的交通方式中，铁路占的比重较大，以城市之间的铁路距离作为城市之间的距离客观合理，具体的数据来源是中国铁路网的统计数据。距离衰减系数 b 取 1.5，重力系数 k 取 1，地区人口总量 Q_i 与 Q_j 数据来源于第六次全国人口普查统计数据。

计算每个省 i 或者直辖市 i 的排名前五的引力向量 S_i^{top5}

$$S_i^{\text{top5}} = [S_{i1}\,S_{i2}\,S_{i3}\,S_{i4}\,S_{i5}]$$

$S_{i1}\,S_{i2}\,S_{i3}\,S_{i4}\,S_{i5}$ 分别为与省 i 或者直辖市 i 之间引力最大的前五个省（市），在这个基础上获得每个省（市）i 对应的吸引力最大的前五个省（市），并认为省（市）i 与这五个省或直辖市之间存在潜在的人口流动关系，建立人口流动网络连线。根据计算得出的省际人口引力 S_{ij}^{G}，进一步计算省（市）i 的综合引力势能 E_i^{G}：

$$E_i^{G} = \sum_{j \neq i} S_{ij}^{G}$$

综合引力势能 E_i^{G} 反映的是省（市）i 人口流动潜力。综合引力势能 E_i^{G} 越大，说明省（市）i 越有可能成为流动人口的大省，既包含了流出人口大省

也包含了流入人口大省。

借助 UCINET 社会网络分析软件，对建立的基于传统重力模型的预期流动人口网络的凝聚子群、聚类系数以及最短距离等网络特征进行分析，考察流动人口网络的整体网特征。通过分析得知，基于传统重力模型的预期流动人口网络存在以下特征：①流动人口网络呈现小世界特征。在流动人口网络中，聚类系数为 0.49，特征长度为 2.855，说明网络中任意两点之间的平均距离位于 2～3 之间，大部分的省及直辖市节点通过 2 个或 3 个中间省(市)建立人口流动联系，平均经过 2.855 距离产生人口流动，省际预期人口流动网络呈现明显的小世界网络特征，有利于全国范围内人力资源的有效利用和共享，对促进经济繁荣具有重要作用。②预期流动人口网络具有明显的凝聚性，存在若干凝聚子群或者社团结构。根据网络分析结果，省际人口流动具有明显的团体性和聚集性。全国范围来看，主要分为八个区域，如图 8.3 所示，包括东三省、华中、华南、中东部沿海、京津冀等地区。

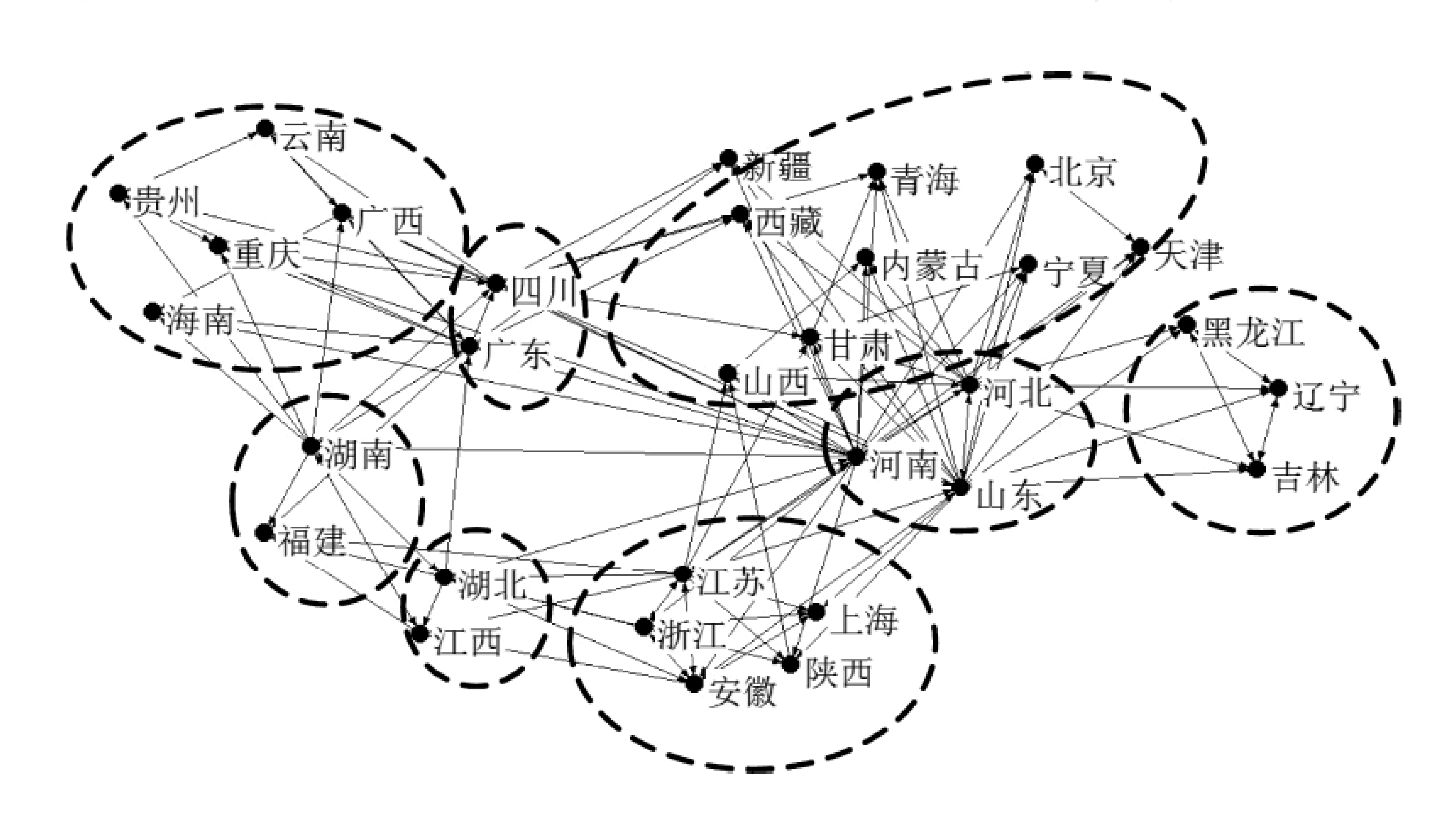


图 8.3　基于重力模型的流动人口网络子群

根据数据计算的综合引力势能结果如表 8.6 所示。

结果显示，根据重力模型计算出的山东、河南、江苏等流动人口大省与实际发生人口流动的主要省(市)结果基本一致，因此基于重力模型的综合引力势能一定程度上反映了一个省(市)的人口流动能力，但是由于引力势能



表 8.6 引力势能前十强的省(市)以及流入、流出人口大省(自治区)

省(市)	引力势能	流入人口大省	流出人口大省
山东	1.286702	广东	安徽
河南	1.278106	上海	河南
江苏	1.250103	北京	四川
安徽	1.118393	浙江	湖南
河北	1.082001	江苏	湖北
湖北	0.914416	福建	江西
湖南	0.837019	天津	广西
广东	0.828007	山东	河北
浙江	0.664251	辽宁	江苏
江西	0.645297	新疆	山东

是一个无向计算指标,所以由此得出的流动人口大省也不具备人口流动的方向性,无法区分是流出人口大省还是流动人口大省。因此需要对流动人口的测度模型进一步改进,构造可以测出方向性的人口流动指标。

3）基于改进重力模型的测度

人口流动受多种因素影响,包括交通成本、经济发展水平、人口以及社会文化因素等。上述重力模型存在两方面不足,一是只考虑了人口和交通两个方面,二是计算得出的测度不具方向性,以此构造的流动人口矩阵是对称矩阵。为了更加准确地度量经济等其他因素对省际人口流动的影响,本案例对重力模型进行了修正,建立了基于交通、经济和人口的重力测度模型,其基本的计算公式如下:

$$R_{ij}^{G} = G\frac{\alpha(P_iP_j)\cdot\beta\left(\frac{Y_j}{Y_i}\right)\cdot\gamma\left(\frac{W_i}{W_j}\right)}{T_{ij}{}^{b}}$$

符号解释:

R_{ij}^{G}:表示省(市)i 的人口流向省(市)j 的预期强度,数值越大,说明从省(市)i 流向省(市)j 的人口数量越多,强度越大。

G：表示重力模型的重力常数，一般取值为 1。

α,β,γ：分别表示人口、经济以及就业情况下不同因素的权重，根据具体实际情况采取不同的权重考虑人口、经济以及就业对人口流动的影响程度。

P_i：表示省(市)i 的人口总量，包括常住人口和外来流动人口，常住人口和流动人口的度量均参照第六次全国人口普查的相关规定。常驻人口＝户口在本辖区人也在本辖区居住＋户口在本辖区之外但在户口登记地半年以上的人＋户口待定(无户口和口袋户口)＋户口在本辖区但离开本辖区半年以下的人。外来流动人口包括离开户口登记地半年以上、来到本辖区不到半年的外来流动人口，不包括户籍人口中离开本辖区半年以上的人。

Y_i：表示省(市)i 的城镇工资总额，指年度内在岗职工工资总额、劳务派遣人员工资总额和其他就业人员工资总额之和。

W_i：表示省(市)i 的城镇登记失业率，具体解释为城镇登记失业人员与城镇单位就业人员(扣除使用的农村劳动力、聘用的离退休人员、港澳台及外方人员)、城镇单位中的不在岗职工、城镇私营业主、个体户主、城镇私营企业和个体就业人员、城镇登记失业人员之和的比。

T_{ij}^b：表示距离阻尼，其中 T_{ij} 为省(市)i 与省(市)j 之间的交通距离，b 是距离衰减系数，衰减系数一般在 1～2 之间。

修正后的模型具有如下特点：

(1) 模型将经济因素、就业影响以及交通成本考虑进去，相对全面地测量了省际之间的人口流动可能性。

(2) 模型测量的人口流动具有方向性。根据计算公式，一般情况下由于各地的经济发展水平及城镇工资总额和失业率不同，计算得出的 $R_{ij}^G \neq R_{ji}^G$，因此根据修正后的重力测度模型计算得出的省际可能存在的人口流动矩阵式是非对称的，符合人口流动的规律。

(3) 模型反映了人口流动的基本规律：经济条件差异越大，距离越近，从一地向另一地的人口流量越大。



5. 基于修正重力模型的预期人口流动网络实证研究

本案例人口、经济以及失业率数据采用国家 2012 年的统计数据以及第六次全国人口普查数据，时间节点为 2012 年，保证了数据的时效与新颖性。由于数据的不完整性，剔除了中国香港、澳门以及台湾，保留了 27 个省（自治区）和 4 个直辖市。省（市）之间的距离统计是按照中国铁路通达性进行计算的，主要基于以下考虑，中国的铁路覆盖面积比较广，城市之间的交通方式中，铁路占的比重较大，以城市之间的铁路距离作为城市之间的距离客观合理，具体的数据来源是中国铁路网的统计数据。为了计算的简便性以及可操作性，省际之间的距离以省会城市之间的距离代替。

1）省际流动人口权利的量化分析

社会网络分析中，省际流动人口网络中节点省市的权利量化指标有中心度和中心势指数，中心度指标是对个体权利的量化分析指标，而中心势指数是对群体权利的量化分析。度数中心度越大，表明相邻的节点数目越多，该节点的地位越核心，在流动人口网络中占据关键地位。预期省际人口流动网络是权重有向网络，其权重的大小与预期人口迁移量的大小是一致的，因此预期省际人口流动网络的中心度分析在一定程度上反映了跨省预期人口流量的核心省（市）情况。

对省际流动人口网络的中心性分析得到表 8.7。计算结果显示，人口流出大省依然集中在中部地区，包括安徽、河南、河北、湖南、湖北等省，东部沿海地区的山东省在人口流出方面比较突出，而西部地区流出人口大省则以山西和陕西为住；在人口流入方面，广东省的中心度占据首位，随后是山东、江苏、北京和上海。根据社会网络分析结果的绘制社会网络图 8.4（图中的节点大小代表节点中心度属性的相对大小）直观显示，未来人口流动依然活跃于中东部地区，呈现两纵的格局：中部狭长区域人口流出为主，东部沿海狭长区域人口流出为主。人口流出区域主要集中于以安徽省为关键地位的中部狭长地域，一方面与中部地区的经济发展水平较低有关，大量的人口寻

求外省的经济效益而发生人口迁出行为；另一方面中部地区的人口数量大，劳动力富裕，是劳动力输出的典型区域。人口迁出区域主要集中在东部沿海地区，其中广东省为主要的迁入大省。从宏观上看，我国的预期人口流动状态是中部流向东部沿海地区，而西部地区的预期人口流动强度（迁出和迁出）相对较弱，但是否发生大规模内部迁移有待进一步的研究。此外通过观察对比预期人口流动网络中各省市节点的出度和入度情况，山东省和安徽省的出度和入度都比较突出，说明这两个省市在人口的迁出和迁入方面都处于网络中相对重要的关键地位，控制着人口流动网络的人力资源流动，因此需要政府加强对这两个省市的人口流动控制与引导，充分发挥网络资源的配置作用。网络的出度和入度中心势分别为6.974%和14.728%，相对很小，说明预期流动人口网络的整体中心性不是很强，网络中存在大量地位相对平等的节点，这与前面分析的人口流动的区域性一致。从图8.4可以看出人口流动强度较大的几个省市包括湖南、湖北、河南、安徽以及河北等地均分布在中部地区，因此对人口流动的相关政策倾斜应以中部地区为主导。

表8.7　省际流动人口网络出度与入度（前十名）

省（自治区）、市	出度	省（自治区）、市	入度
安徽	2372894.75	广东	4124329
河北	1916444.25	山东	2978315
湖南	1698649.375	江苏	2695266
江西	1480265.125	北京	2295092
湖北	1449730.5	上海	2190427
山西	1244735.25	浙江	1790209
河南	1173993.625	河北	1071571
广西	1088378.125	湖北	888464.3
陕西	932560.375	四川	775558.1
山东	884350.563	安徽	771666.9

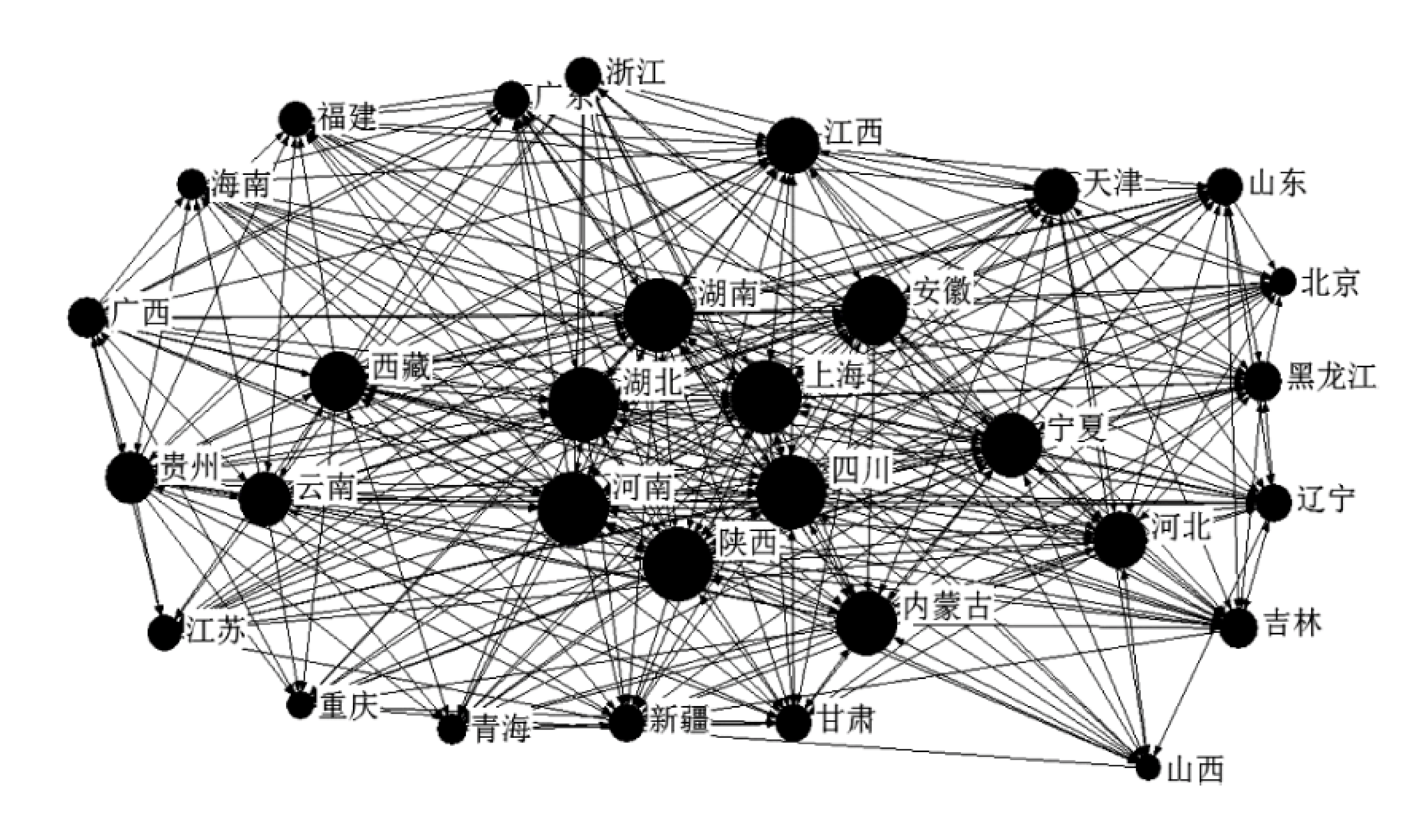


图 8.4　基于中心度属性的人口流动的社会网络属性图

（节点的大小代表节点中心度属性的相对大小）



2）省市流动人口网络的关联性和小世界分析

社会网络中，关联性测量的是网络中行动者之间的关联特征，关联度低的网络权利集中，信息集中，网络中的行动者地位不平等，并且网络极易受到个别节点的影响，具有一定的分派结构；而关联度高的网络，权利分散，信息分散，网络中行动者之间的地位是平等的，不容易受到个别点的影响，网络结构是均匀的。根据修正的重力模型建立的预期省际流动人口网络是完备网络，网络中的所有节点均互相连接。为了进一步探讨流动网络的关联性和小世界特征，对流动网络进行进一步处理：保留每个与每个节点连接强度在前五之内的连线，剔除其余连线，得到新的简化的高强度预期省际人口流动网络，以便更好地分析网络的拓扑结构。计算结果如表 8.8 所示。

表 8.8　省际流动人口网络的关联性和小世界指标

网络指标	具体数值	网络指标	具体数值
关联度(Connectedness)	0.4914	聚类系数(Clustering coefficient)	0.624
效率(Efficiency)	0.9016	距离(Distance)	1.548
最近上限(LUB)	1.0000		

在预期省际流动人口网络中，关联度指数值为 0.4914，说明大规模的人口流动网络中节点的可达性在 50%左右。网络的效率为 0.9016，理论上来说在已知网络中所包含的成分数确定的情况下，有近 90%的网络连线是多余的，但是多余的网络联系反映了大规模预期人口流动网络的紧密联系性。最近上限为 1，说明大规模预期人口流动网络的任何省节点之间都具有共同的邻居，可以通过共同的邻居建立起联系，加快人口的流动和人力资源的共享和传递。网络的聚类系数和特征途径长度是小世界的测量指标，聚类系数是局部网络结构的测度指标而特征路径长度是网络整体性质的测度指标，小世界网络具有相对较小的距离和相对较大的聚类系数。在预期省际流动人口网络中，聚类系数为 0.624，特征长度为 1.548，说明网络中任意两点之间的平均距离位于 1～2 之间，大部分的省点通过 1～2 个中间省建立联系，平均经过 1.548 距离建立合作关系，预期省际流动人口网络呈现明显的小世界网络特征。小世界特征的省际流动人口网络有利于网络中人口的流动和人力资源的传播，对省市之间的人力合作和经济发展具有重大意义。

3）预期省际流动人口网络的结构洞与凝聚性分析

在社会网络中，结构洞指的是两个行动者之间的非冗余的联系。根据伯特理论，结构洞能够为其占据者获取信息利益和控制利益提供机会，从而比网络中其他位置成员更具有竞争优势。结构洞的衡量指标分为四个方面：有效规模、效率、限制度以及等级度。由于限制度是衡量结构洞的最重要的指标，因此本案例分析中重点考虑限制度。采用 UCINET 软件对预期省际大规模流动人口网络进行结构洞分析，得到指标数值如表 8.9 所示。

表 8.9　省际流动人口网络的结构洞指标

省（自治区）、市	有效规模	效率	限制度	等级度
河北	11.694	0.487	0.149	0.050
河南	14.924	0.497	0.140	0.057
陕西	15.054	0.502	0.140	0.057
山西	3.917	0.301	0.181	0.049

续表

省(自治区)、市	有效规模	效率	限制度	等级度
山东	5.891	0.347	0.175	0.036
甘肃	6.104	0.359	0.167	0.034
辽宁	5.895	0.347	0.170	0.019
吉林	6.833	0.380	0.164	0.024
黑龙江	7.326	0.407	0.168	0.034
云南	9.838	0.428	0.154	0.049
贵州	9.426	0.428	0.157	0.056
广东	5.522	0.325	0.170	0.046
福建	4.053	0.253	0.179	0.036
海南	3.281	0.219	0.179	0.018
四川	15.295	0.510	0.140	0.058
湖北	14.878	0.496	0.141	0.060
湖南	15.148	0.505	0.141	0.059
江西	10.868	0.453	0.150	0.049
安徽	13.750	0.491	0.143	0.056
江苏	4.474	0.280	0.170	0.039
浙江	4.684	0.276	0.168	0.029
青海	4.184	0.279	0.177	0.029
新疆	5.583	0.328	0.165	0.019
内蒙古	13.385	0.496	0.142	0.053
宁夏	14.662	0.543	0.136	0.054
西藏	11.236	0.449	0.149	0.062
广西	6.217	0.345	0.168	0.042
北京	3.733	0.267	0.177	0.019
天津	8.903	0.424	0.157	0.046
重庆	3.611	0.258	0.184	0.030
上海	15.278	0.509	0.140	0.059

根据结果显示，结构洞的限制度较大的省市有广东、福建、海南、四川、江苏、青海、北京和重庆等，其中西部较偏远的省市占据了近50%，说明西部城市在预期的省际大规模流动人口网络中运用结构洞的能力较弱，受限制性比较强，因此西部偏远省市对其他省市的人口流动依赖性比较强。为了降低这种依赖性，应鼓励中东部人口流动密集的省市与西部地区的人口互动，鼓励人力资源向西部迁移，加快西部地区的发展。与此同时，北京、江苏、广东的人口迁入发达省市的受限制性也比较强，这是因为全国预期大部分地区的人口流向这三个典型地区的数量都比较强，因此各个省市累积的限制度就比较高。从限制度矩阵看，对北京、广东和江苏限制性比较强的省市主要分为两类，一类是西部偏远省市，一类是中部人口流动大省，因此需要降低北京、广东和江苏的限制性，也需要加强中部人口流动大省和西部偏远城市的人口流动强度。

网络的凝聚性分析是通过对网络中心行动者子集的特征来刻画与研究社会群体，采用多种网络属性对群体的凝聚性进行量化处理，比如派系、k丛、k核以及块分析等，其中派系主要是建立在互惠性基础上的凝聚子群，k丛、k核是建立在点度上的凝聚子群，块是建立在子群内外关系基础上的凝聚子群。采用UCINET软件对预期省际人口流动网进行聚类分析，计算出省际流动人口网络存在的8个位置(块)，得到8个块。表8.10列出了每个网络子群的分类情况。

表8.10　省际流动人口网络子群的分类情况

子群	成　　员	子群	成　　员
子群1	河北、内蒙古、山西、陕西、天津、宁夏	子群5	河南
子群2	吉林、辽宁、黑龙江	子群6	山东、北京
子群3	甘肃、新疆、青海、安徽、西藏	子群7	广东
子群4	福建、广西、湖南、江西、贵州、湖北、重庆、四川、海南、云南	子群8	浙江、江苏、上海

根据凝聚子群分类情况,发现生成的人口流动子群与地域位置关系密不可分,预期省际人口流动主要分为8个区域(见图8.5):东三省地区、华北地区、西部地区、长三角地区、珠三角地区、京津冀地区、东南地区、河南地区,具有一定的地域分布特性,其中河南省和广东省各自分别成为一个社区子群。结合预期人口流动网络的中心性分析,可以推测,在人口流出方面,河南省作为人口资源大省控制着整个预期省际人口流动的网络资源,应予以重点关注。广东省作为典型的人口流入大省在接纳外来人口方面居于网络中的关键位置,在人口流动网络中发挥重要的核心引领作用。根据区域控制以引领网络中的人口合理流动,对于省际人力资源的共享和省际经济的发展具有重要的作用。

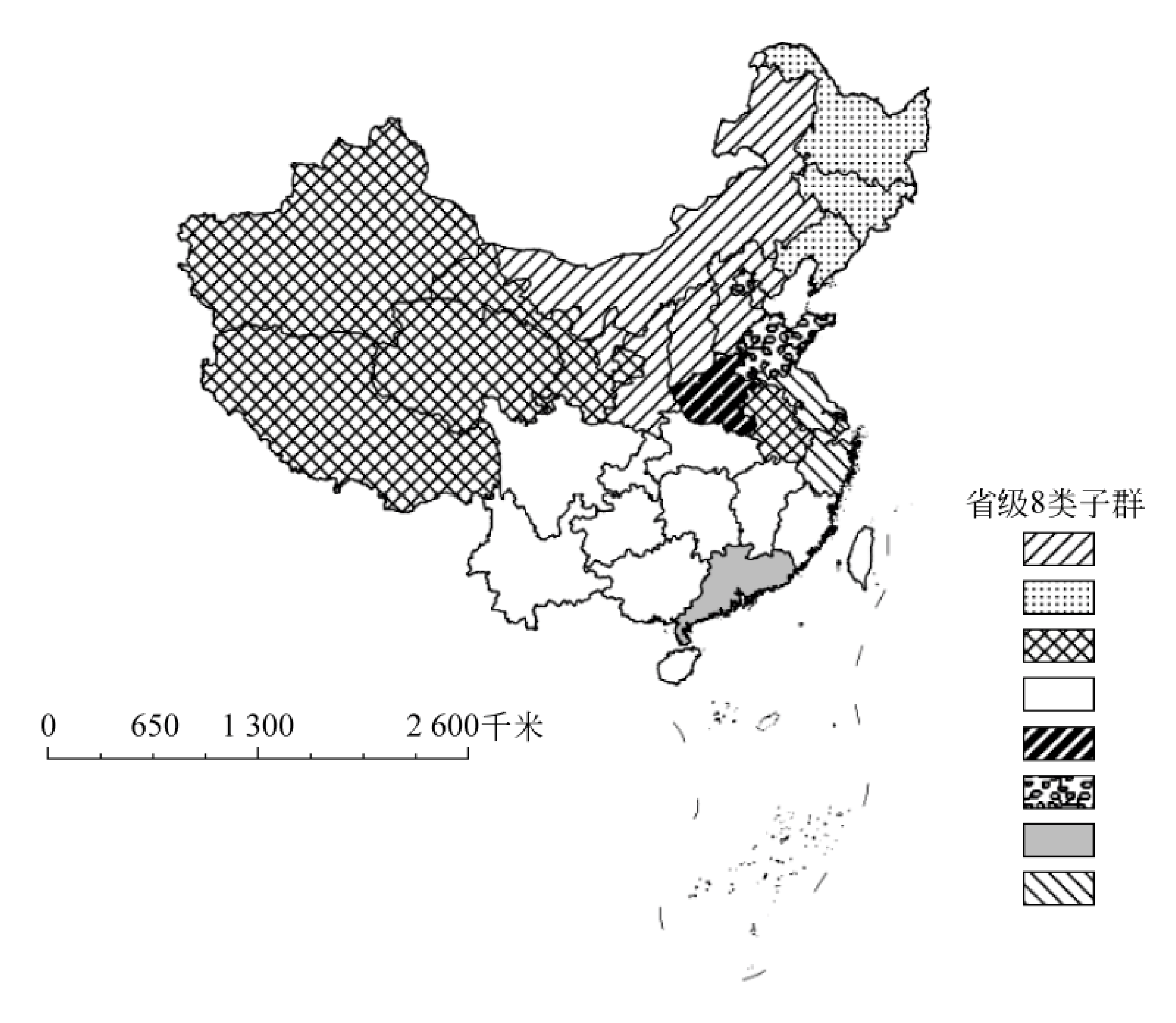


图 8.5　人口流动的省群分布

整个网络的密度值经过计算为26585.3750,将计算得出的密度矩阵表中大于26585.3750的值都修改为1,小于26585.3750的值都修改为0,得到的子群密度矩阵如表8.11所示。

表 8.11 子群密度矩阵

	子群 1	子群 2	子群 3	子群 4	子群 5	子群 6	子群 7	子群 8
子群 1	0	0	0	0	1	1	1	1
子群 2	0	1	0	0	1	1	1	1
子群 3	0	0	0	0	1	1	1	1
子群 4	0	0	0	0	1	1	1	1
子群 5	1	0	0	0	0	1	1	1
子群 6	0	0	0	0	1	1	1	1
子群 7	0	0	0	0	0	0	0	0
子群 8	0	0	0	0	1	1	1	1

上述分析得到的是一个非对称矩阵，由此得到如下结论：预期省际流动人口网络分为 8 个子群，部分子群存之间存在人口流动关系，帮派性相对较为明显。人口流动关系比较频繁地出现在子群 5、子群 6、子群 7 以及子群 8 之间，东三省地区代表的子群 2 内部存在显著的人口流动关系。广东为代表的子群 7 与其他子群之间只存在人口迁入的单向关系，子群 1、子群 2、子群 3 以及子群 4 之间不存在明显的人口流动关系。以子群为节点，子群内部及子群之间的人口流动关系为边得到简化图，见图 8.6。



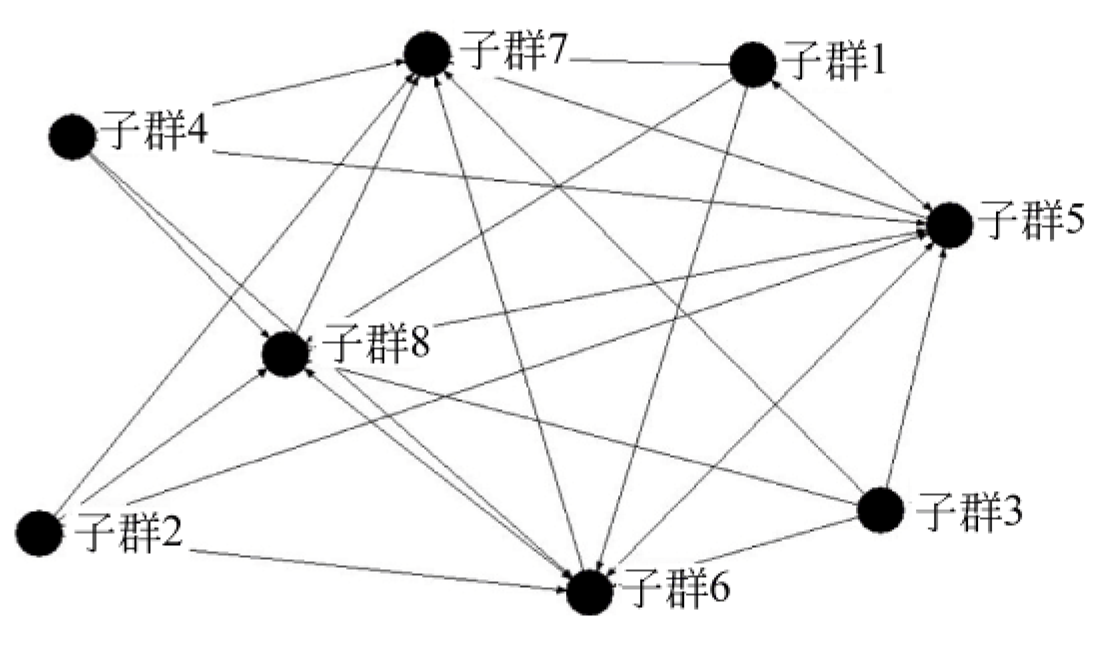


图 8.6 子群关系简化图

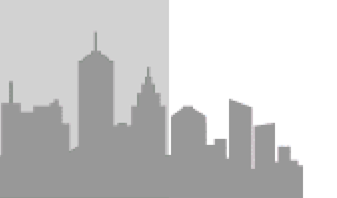
本案例将复杂的理论与社会网络分析方法引入到省际流动人口研究中，根据实际的人口流动建立复杂网络，分析大规模流动人口的网络拓扑结构以及以人口流动量为权重的网络特征和流动人口的空间分布特征。通过

对比分析基于交通成本和重力模型测度的预期人口流动与实际人口流动的分布，建立了基于交通、经济和人口流动的改进重力模型，测度省际之间的人口流动强度，以省份为节点，省际预期的人口流动强度为连接边，构建省际预期流动人口有向网络，借助复杂网络分析方法和社会网络分析工具UCINET，探讨了网络的中心性、凝聚性、无标度性以及小世界特征。本案例研究有助于政府管理部门为流动人口管理以及产业布局提供科学决策。

通过对省际实际人口流动网络以及省际预期人口流动网络的分析，得到以下几点结论：

（1）大规模实际人口流动网络呈现小世界特征，无标度特征不明显，度分布指数较小，流动人口网络的异质性比较高，网络中不同度值的节点分布均匀。

（2）地区较高经济发展水平和较大人口总量产生对外来人口较强的吸引力。实际流出人口省份主要集中在以安徽和河南为代表的中部地区，人口流入大省主要集中在北上广以及东部沿海等经济较为发达的省份，其中新疆作为国家西部开发的重点区域，在近年来的人口流入省份中占有重要地位。



（3）基于交通、经济和人口的改进重力模型测度的预期省际人口流动网络分析得出，预期流动人口网络的整体中心性不是很强，网络中存在大量地位相对平等的节点，未来人口流动依然活跃于中东部地区，呈现两纵的格局：中部狭长区域人口流出为主，东部沿海狭长区域人口流入为主，其中网络指标显示，未来山东省和安徽省在人口的迁出和迁入方面都处于网络中相对重要的关键地位，控制着人口流动网络的人力资源流动。

（4）西部城市在预期的省际大规模流动人口网络中运用结构洞的能力较弱，受限制性比较大，对其他省市的人口流动依赖性比较强。全网分为八个子群社团，具有明显的地域分布特征，受空间以及人口经济因素的影响，部分社团之间联系频繁，帮派性相对较为明显，网络具有较高的聚类系数和较小的平均距离，小世界特征明显，具有较强的局部聚集性和整体连通性。

8.2 基于凝聚子群的信息协同网络结构分析

8.2.1 关联关系矩阵构建

1. 信息组织关联关系矩阵

对集合 $A=\{a_1,a_2,\cdots,a_n\}$，为了表示方便，将信息组织 $a_s(s=1,2,\cdots,n)$ 简记为信息组织 s。令 $a_{ij}(i,j=1,2,\cdots,n)$ 表示信息组织 i 向信息组织 j 共享的信息数量（类）或信息流量（次/天），则信息组织的关系矩阵 $\boldsymbol{P}_A$ 表示为：

$$\boldsymbol{P}_A=\begin{vmatrix} a_{11}, & a_{12}, & \cdots, & a_{1n} \\ a_{21}, & a_{22}, & \cdots, & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1}, & a_{n2}, & \cdots, & a_{nn} \end{vmatrix}$$

一般情况下，$a_{ij}\neq a_{ji}$，且 $a_{ii}=M$（M 表示足够大的正数）。



2. 组织-信息关联关系矩阵

令 $B=\{b_1,b_2,\cdots,b_w\}$ 表示 w 类信息的集合，为了表示方便，将信息 $b_t(t=1,2,\cdots,w)$ 简记为信息 t。

对信息 t，令 $\overline{\boldsymbol{Q}_t}=(c_{1t},c_{2t},\cdots,c_{nt})^{\mathrm{T}}$，其中

$$c_{st}=\begin{cases} 0, & \text{组织 } s \text{ 是信息 } t \text{ 的需求方} \\ 1, & \text{组织 } s \text{ 不是信息 } t \text{ 的需求方} \end{cases} \quad (s=1,2,\cdots,n;t=1,2,\cdots,w) \tag{8.1}$$

则组织-信息关联关系矩阵

$$\boldsymbol{P}_{AB}=(\boldsymbol{Q}_1,\boldsymbol{Q}_2,\cdots,\boldsymbol{Q}_w)=\begin{vmatrix} c_{11}, & c_{12}, & \cdots, & c_{1w} \\ c_{21}, & c_{22}, & \cdots, & c_{2w} \\ \vdots & \vdots & & \vdots \\ c_{n1}, & c_{n2}, & \cdots, & c_{nw} \end{vmatrix}$$

一般情况下，一个信息组织至少有一类信息，因此 $w \geqslant n$。

3. 信息关联关系矩阵

对信息 i 和信息 j，令 b_{ij} 表示信息 i 与信息 j 的共同需求方数量，有

$$b_{ij}=\sum_{k=1}^{n}(c_{ki}c_{kj}),\quad (i,j=1,2,\cdots,w) \tag{8.2}$$

b_{ij} 越大，说明信息 i 与信息 j 的关联程度越高。

构建信息的关联关系矩阵

$$\boldsymbol{P_B}=\begin{vmatrix} b_{11}, & b_{12}, & \cdots, & b_{1w} \\ b_{21}, & b_{22}, & \cdots, & b_{2w} \\ \vdots & \vdots & & \vdots \\ b_{w1}, & b_{w2}, & \cdots, & b_{ww} \end{vmatrix}$$

矩阵 $\boldsymbol{P_B}$ 满足关系：①$b_{ij}=b_{ji}$；②$0\leqslant b_{ij}\leqslant n-2$。

令 $b'_{ij}=\dfrac{b_{ij}}{n}$ 表示信息 i 和信息 j 之间的距离，将关系矩阵 $\boldsymbol{P_B}$ 转换为距离矩阵 $\boldsymbol{P'_B}$

$$\boldsymbol{P'_B}=\begin{vmatrix} b'_{11}, & b'_{12}, & \cdots, & b'_{1w} \\ b'_{21}, & b'_{22}, & \cdots, & b'_{2w} \\ \vdots & \vdots & & \vdots \\ b'_{w1}, & b'_{w2}, & \cdots, & b'_{ww} \end{vmatrix}$$

8.2.2 信息协同网络的凝聚子群分析

凝聚子群(cohesive subgroup)包括派系(cliques)、n-派系、n-宗派、k-丛、k-核等多种类型。一个派系是一个行动者的子集合，子集合中行动者之间的联系相对比较紧密。理解网络结构及个体嵌入性的一个重要途径就是对子结构(或子群、派系等)的分析。

对信息协同网络的凝聚子群分析主要建立在“子群内外关系”的基础上，通过构建块模型(block model)划分子群，进行小群体的量化研究。块模

型的构建主要通过 CONCOR 聚类方法实现。McQuitty 于 1968 年发现了相关系数矩阵迭代的收敛性，Breiger 等于 1975 年将这种迭代方法命名为 CONCOR 算法。近年来人们将其用于聚类，并称为 CONCOR 聚类，主要应用于将网络结构分成不同的子群或块，得到树型图，发现网络中的结构特点，然后对结果进行分析理解。

CONCOR 聚类的基本思想与其他方法不同。这种聚类方法对数据没有特别要求，任何数据都可使用，主要借助于矩阵的 CONCOR 变换。

定义 1　CONCOR 变换：若 $\boldsymbol{A}=(a_{ij})_{n\times m}$ 是一个 n 行 m 列的矩阵，则以 $\boldsymbol{A}$ 的 i 行与 j 行的相关系数

$$b_{ij}=\frac{\sum_{k=1}^{m}(a_{ik}-\bar{\delta}_i)(a_{jk}-\bar{\delta}_j)}{\sqrt{\sum_{k=1}^{m}(a_{ik}-\bar{\delta}_i)^2}\sqrt{\sum_{k=1}^{m}(a_{jk}-\bar{\delta}_j)^2}} \tag{8.3}$$

为矩阵 $\boldsymbol{B}=(b_{ij})_{n\times n}$ 的 i 行 j 列元素，这里 $\bar{\delta}_i$，$\bar{\delta}_j$ 分别是 $\boldsymbol{A}$ 的第 i 行及第 j 行的平均值，矩阵 $\boldsymbol{B}$ 称为 $\boldsymbol{A}$ 的 CONCOR 变换，记作 $\boldsymbol{B}=\text{CONCOR}(\boldsymbol{A})$。

利用 CONCOR 变换进行 CONCOR 聚类的过程如下：

（1）构建 $n\times m$ 矩阵 $\boldsymbol{V}=(v_{ij})_{n\times m}$，其中 $v_{ij}(1\leqslant i\leqslant n, 1\leqslant j\leqslant m)$ 为第 i 个对象在第 j 个属性上的取值。

（2）令 $\boldsymbol{C}_1=\text{CONCOR}(\boldsymbol{V}),\cdots,\boldsymbol{C}_n=\text{CONCOR}(\boldsymbol{C}_{n-1}),\cdots$，已证明迭代序列 $\boldsymbol{C}_1,\boldsymbol{C}_2,\cdots,\boldsymbol{C}_n,\cdots$ 一定在有限次收敛。迭代过程中，除 $\boldsymbol{V}$ 是 $n\times m$ 阶矩阵之外，$\boldsymbol{C}_1,\boldsymbol{C}_2,\cdots,\boldsymbol{C}_n,\cdots$ 都是 $n\times n$ 阶矩阵。

（3）将这个极限矩阵中相等的各行所对应的对象作为一类，就是对 n 个对象的 CONCOR 聚类。

（4）如果聚类结果不能满足用户的要求（截止条件，如类数大于某个值或各类的对象个数小于某个值等），则对每一类对应的子阵继续进行 CONCOR 聚类，将每一类进一步细分；不断循环该过程，直到满足截止条件为止。

对信息组织关系矩阵 $\boldsymbol{P}_A$ 和信息距离矩阵 $\boldsymbol{P}'_B$，由于不同的子群中存在一定的交叉关系，在 CONCOR 聚类的基础上，利用 UCINET 中的凝聚子群相关算法进行进一步的派系分析，凝聚成若干个存在交叉关系的子群(或多级子群)。其中，对信息协同网络中的 w 类信息，令 $w=w_1+w_2$，w_1 表示与其他信息之间没有共同需求方或只有 1 个共同需求方的信息数量，w_2 表示与其他信息之间存在两个及以上共同需求方的信息数量。存在两个及以上共同需求方的多源信息之间具有信息融合的可能性，是网络结构分析的重点。

8.2.3 多源信息协同模式的纵向优化策略

从信息协同网络的纵向结构上，根据多源信息协同的关系测度分析，在多级凝聚子群的基础上，将 PCP 模式转变为以领域为中心的二级(或多级) P2N 模式。一方面，通过改变信息流向，将控制端前移，利于具有共性需求的信息协同流程管控；另一方面，对来自不同信息组织、具有相同需求方的信息预先进行整合，从而优化信息协同的路径，降低信息协同网络的复杂度，减少需求端信息重复整合的次数，提高信息流转的效率。二级协同网络内部的模式优化策略和过程与一级协同网络相同。



应用案例 7　智慧城市多源信息协同的评价与优化(下)

在应用案例 5 中介绍了某智慧城市多源信息协同的差异化水平评价与模式优化策略。下面继续对该市的多源信息协同网络进行分析。

对 24 个信息组织建立关系矩阵 $\boldsymbol{P}_A$，得到有向赋权网络图(见图 8.7)。其中，节点大小表示信息组织的协同信息量，边的属性值表示信息流量，边的方向表示信息流向。

对 155 类协同信息建立距离矩阵 $\boldsymbol{P}'_B$，其中共同需求方 <2 的信息 74 类，共同需求方 $\geqslant 2$ 的信息 81 类。对存在两个及以上共同需求方的协同信息进行分析，子群分布如图 8.8 所示。

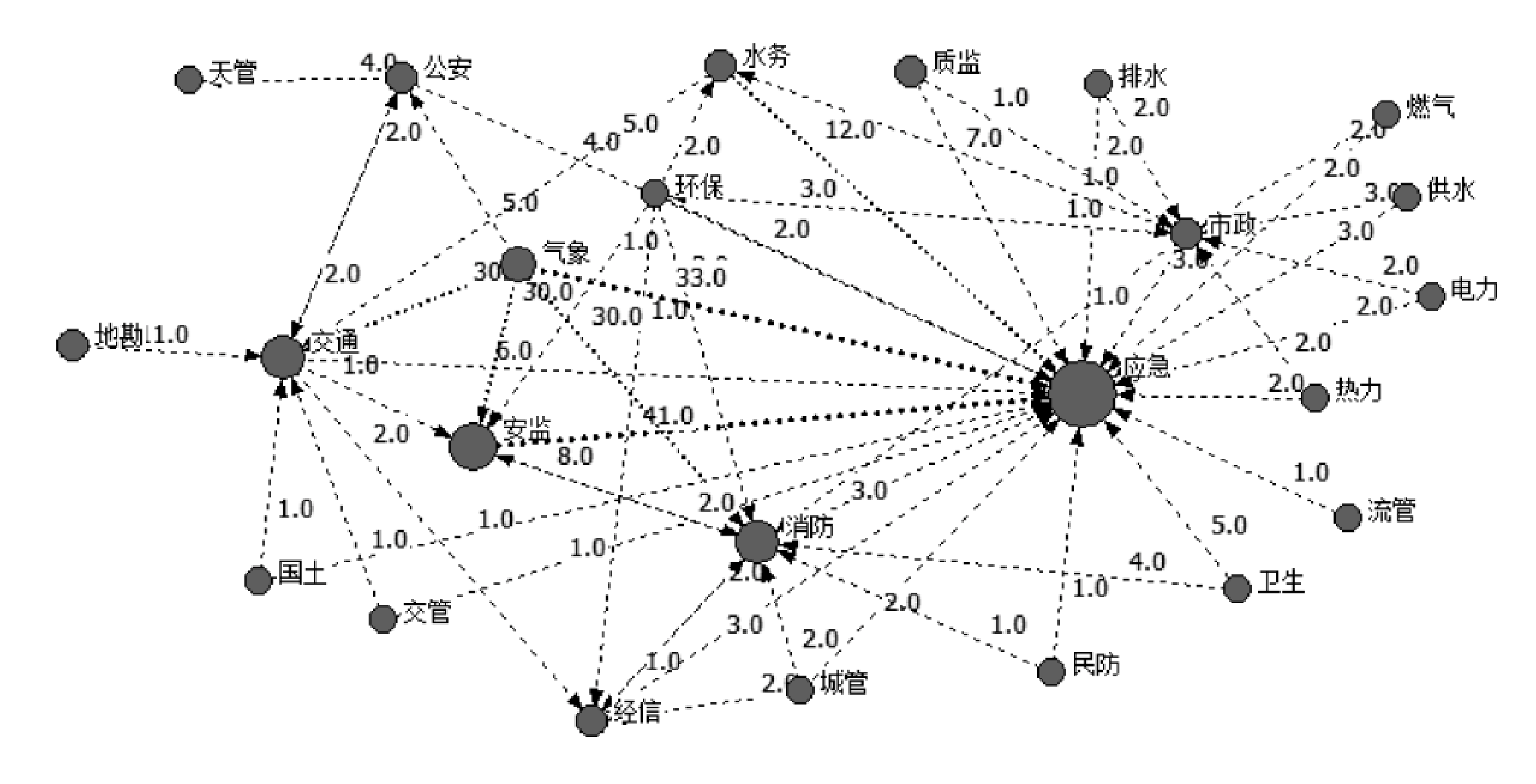


图 8.7 某智慧城市基础运行领域的信息组织协同网络结构

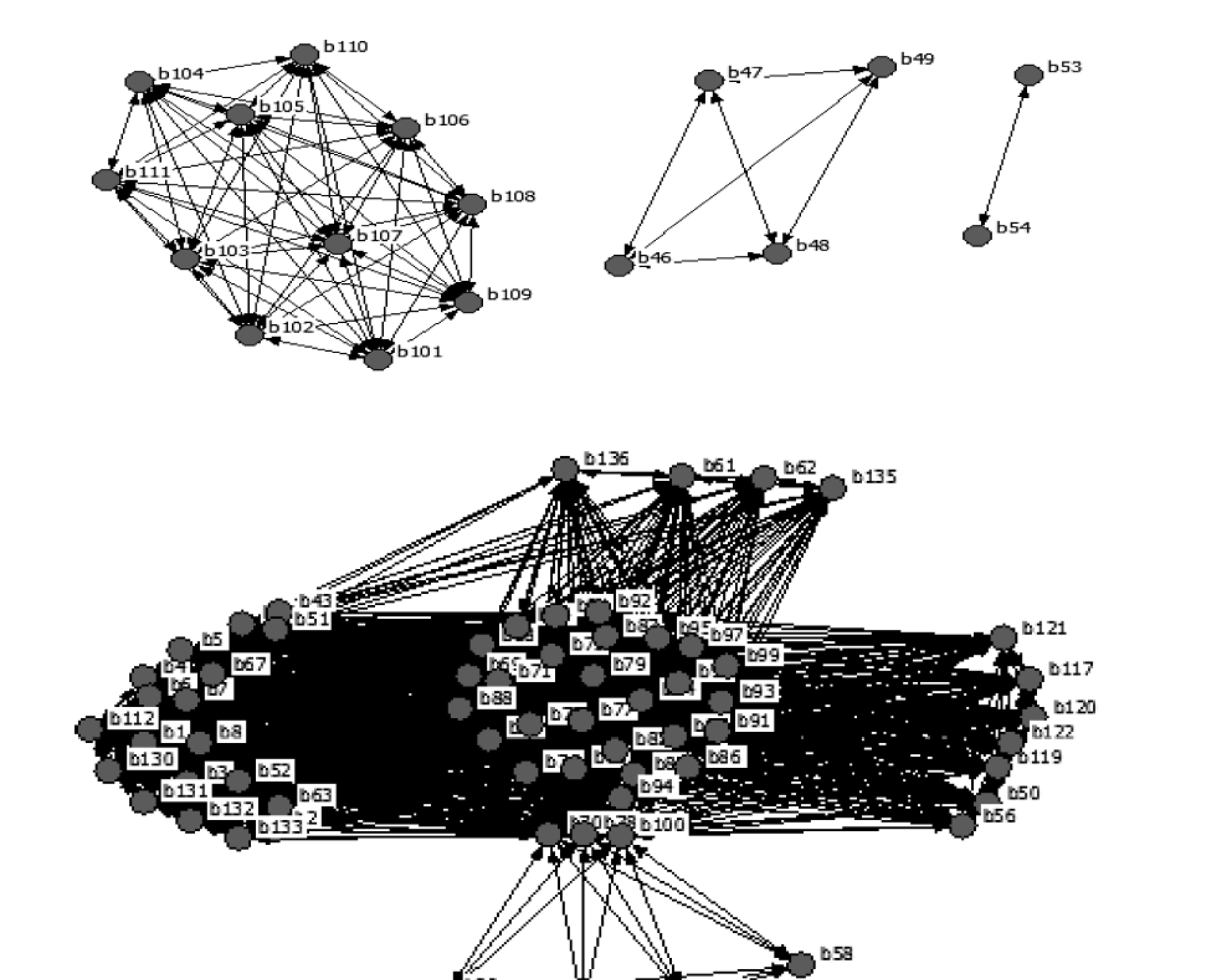


图 8.8 某智慧城市基础运行领域的协同信息子群分布

由图 8.8 可知,81 类信息分成 4 个一级子群(见表 8.12)。

表 8.12　协同信息的一级子群分布

一 级 子 群	子群 1	子群 2	子群 3	子群 4
包含的信息类数	4	2	11	64

其中一级子群 4 中又包含 5 个二级子群(见表 8.13)。

表 8.13　一级子群 4 的二级子群分布

二 级 子 群	子群 1	子群 2	子群 3	子群 4	子群 5
包含的信息类数	4	4	7	19	30

对矩阵 $\boldsymbol{P}_A$ 和矩阵 $\boldsymbol{P}_B'$ 进行凝聚子群中的 CONCOR 聚类和派系分析,发现信息组织在信息协同方面具有明显的领域凝聚性。从总体结构来看,主要分为 7 个子群(见图 8.9)。

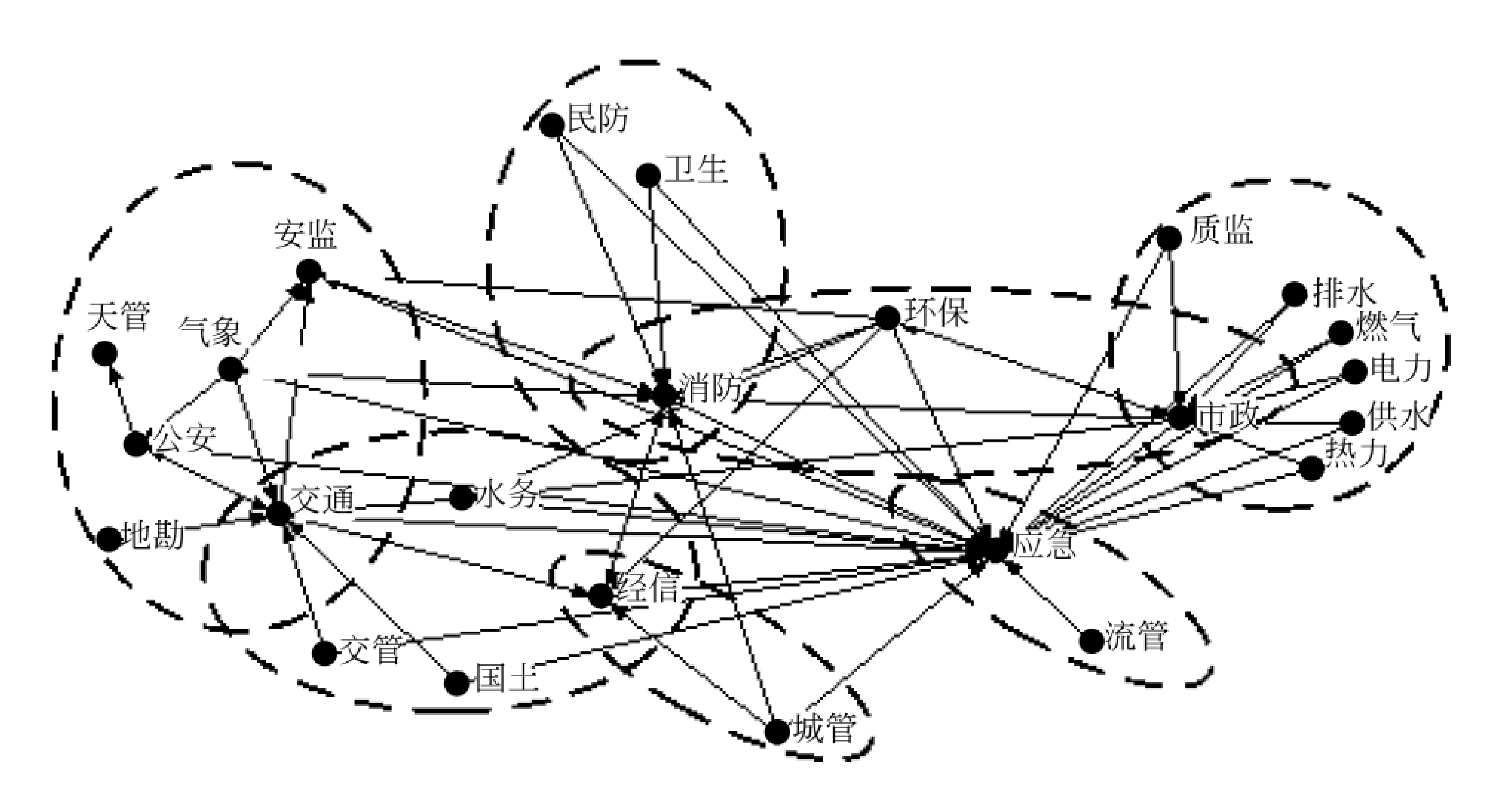


图 8.9　某智慧城市基础运行领域的信息组织协同子群分布

由结果分析可知,城市生命线管理(市政、质监、供水、排水、电力、燃气、热力)、流动人口管理(应急、流管)、城管执法(经信、城管)、安防急救(消防、卫生、民防)、危险化学品管理(安全监管、气象、交通、公安等)、应急防汛(交

通、水务、国土等)、市政管理(市政、环保、消防)7 个领域的信息组织具有较强的凝聚性,信息协同应用程度较高。

此外,交通与气象、交通与水务、市政与环保等领域之间具有较强的凝聚性(信息关联性),分别对应当前大型城市基础运行领域实际应用中的几个热点问题,包括极端天气下的交通保畅、汛期路面积水造成的交通拥堵、空气质量与环境噪声监管等。

根据信息组织的协同子群分布和信息的二级子群分布结果,将传统的 PCP 模式转变为 7 个以领域牵头部门为二级中心的 P2N 模式。对于凝聚性较强的信息组织间的多源信息,根据需求进行整合形成综合信息(如路面积水点实时数据与交通流量图层叠加),减少同类协同信息的重复传输和需求端的重复整合。通过改变信息的流向,从纵向结构上实现城市系统下的信息协同网络的优化,提高信息协同的灵活度和协同应用的实效性。

附录 A　中国和美国政府的大数据战略比较



1. 中美政府大数据战略的启动

2012 年 3 月 29 日，美国奥巴马政府宣布启动“**大数据研究与开发计划**”。该计划由白宫科技政策办公室与美国联邦政府的美国国家科学基金（NSF）、美国国家卫生研究院（NIH）、美国能源部（DOE）、美国国防部（DOD）、美国国防部高级研究计划局（DARPA）、美国地质勘探局（USGS）六个部门共同制定，投入超过两亿美元的资金，大力推动和改善与大数据相关的收集、组织和分析工具及技术，并主要用于对海量数据的访问、组织与信息提取。

2015 年 8 月 19 日，中国国务院总理李克强主持召开国务院常务会议，通过《关于促进大数据发展的行动纲要》。9 月 5 日，《**国务院关于印发促进大数据发展行动纲要的通知**》（国发〔2015〕50 号）正式发布。这是到目前为止中国促进大数据发展的第一份权威性、系统性文件，从国家大数据发展战略全局的高度，提出了中国大数据发展的顶层设计，是指导中国未来大数据发展的纲领性文件，核心是推动各部门、各地区、各行业、各领域的数据资源共享开放。

2. 中美政府大数据战略的宗旨和目标比较

美国“大数据研究与开发计划”的主要宗旨有三个方面：一是推动最新的核心技术研发，以用于海量数据的收集、存储、保护、管理、分析和共享；二是充分利用所研发的核心技术加快科技与工程领域的研发速度，加强国家安全，同时转变教育和学习模式；三是培养更多大数据技术开发与使用方面的专业人才。

中国《关于促进大数据发展的行动纲要》的主要目标有五个方面：一是打造精准治理、多方协作的社会治理新模式；二是建立运行平稳、安全高效

的经济运行新机制;三是构建以人为本、惠及全民的民生服务新体系;四是开启大众创业、万众创新的创新驱动新格局;五是培育高端智能、新兴繁荣的产业发展新生态。

3. 中美政府大数据战略的项目布局对比

美国“大数据研究与开发计划”主要包括以下项目:美国国家科学基金会和美国国家卫生研究院的“推进大数据科学和工程的核心方法及技术”、美国国防部的“利用数据支持决策”、美国国家卫生研究院的“千人基因组计划的数据在亚马逊云上免费开放”、美国能源部的“通过高级计算技术加速科学发现”、美国地质勘探局的“地球科学领域的大数据”、国土安全部的“卓越研究中心和可视化数据分析”,以及退伍军人管理部、卫生和人类服务部、国家档案和记录管理部、国家航空和航天局、国家人文基金会、国家安全局等部门的一系列具体项目。

中国《关于促进大数据发展的行动纲要》主要包括十项具体工程,其中**“政府大数据”**工程4项(政府数据资源共享开放工程、国家大数据资源统筹发展工程、政府治理大数据工程、公共服务大数据工程),**“大数据产业”**工程5项(工业和新兴产业大数据工程、现代农业大数据工程、万众创新大数据工程、大数据关键技术及产品研发与产业化工程、大数据产业支撑能力提升工程)及网络和大数据安全保障工程。

4. 美国政府大数据战略的主要项目介绍

1) 美国国家科学基金会和美国国家卫生研究院——推进大数据科学和工程的核心方法及技术

美国国家科学基金会和美国国家卫生研究院将对大数据进行联合招标,旨在改进核心科学与技术手段,提高从各种大型数据集中提取重要信息并对其进行有效管理、分析和可视化的能力,加速科技成果的产生,并带领国家进入一些全新的、以往不可企及的研究领域。其中,卫生研究院对于与健康与疾病相关的数据集尤其感兴趣,包括影像、分子、细胞、电生理学、化

学、行为、流行病学、临床医学数据集。国家科学基金会除了为大数据招标提供资金维持其关注的基础研究外，还正在实施一个全面的、长期的战略，包括开发新的方法，以便更有效地从数据中进行知识获取；相关基础设施投资，用于大数据科研团体的管理、组织和数据提供等；研究新的教育和人才培养方法。采取的具体举措包括：

(1) 鼓励研究型大学设立跨学科的研究生专业课程，以培养新一代数据科学家和工程师人才。

(2) 向加利福尼亚大学伯克利分校的计算开发(Expeditions in Computing)项目投资 1000 万美元，此项目计划集成三种强大的数据转信息方法，包括机器学习、云计算和众包。

(3) 为“EarthCube”提供第一阶段的资金支持——该系统将允许地球学家获取、分析和共享与地球相关的信息。

(4) 向一个研究培训小组发放 200 万美元的奖金，用于支持一项大学生培训计划，教授他们如何利用图形和可视化工具解析复杂数据。



(5) 为一个由统计学家和生物学家组成的专业研究团体提供 140 万美元的研发资金，资助他们研究蛋白质结构和生物途径。

(6) 召集各个学科和领域的研究人员，共同探讨如何利用大数据转变教育与学习模式。

2) 美国国防部——利用数据支持决策

美国国防部“在大数据上压下了巨大赌注”，每年将投资 2.5 亿美元左右(其中 6000 万美元用于支持新的研究项目)在各个军事部门开展一系列研究计划，旨在：

(1) 以创新方式使用海量数据，通过感知、认知和决策支持的结合，建立真正能够独立完成操控并做出决策的自治式系统。

(2) 提高作战人员和分析人员的环境与状况感知能力，增强对任务和流程的支持。国防部的目标是将分析人员从任意语言文字资料中提取信息的能力提高 100 倍，同时希望他们观察到的目标、活动和事件的数量也获得相

同幅度的提升。

具体的项目包括：

（1）多尺度异常检测项目。旨在解决大规模数据集的异常检测和特征化。目前多尺度异常检测应用程序能够进行内部威胁检测，以及在日常网络活动环境中检测单独的异常行动。

（2）网络内部威胁计划。旨在开发新的方法来检测军事计算机网络与网络间谍活动。

（3）洞悉计划。主要解决现有情报、监视和侦察系统的不足，进行自动化和人机集成推理，使得能够提前对时间敏感的更大潜在威胁进行分析。该计划旨在开发出资源管理系统，通过分析图像和非图像的传感器信息和其他来源的信息，进行网络威胁的自动识别和非常规的战争行为。

（4）阅读机项目。旨在实现人工智能的应用和发展学习系统的过程中对自然文本进行知识插入，而不是依靠昂贵和费时的知识表示目前的进程，并需要专家和相关知识工程师所给出的语义表示信息。



（5）想象力项目。旨在为机器建立视觉的智能。传统的机器视觉研究的对象选取广泛的物体来描述一个场景的属性名词，而想象力项目旨在增加在这些场景的动作认识和推理需要的知觉认知基础。

（6）使命导向的高适应性云项目。通过技术进行检测、诊断并对攻击做出响应，有效地建立“社区卫生服务系统”的云，以解决云计算固有的安全挑战。该方案还旨在开发新技术，使云应用和基础设施在受到攻击时能够继续运行。

（7）对加密数据的编程计算的研究工作。旨在开发实用的方法、相关现代化计算编程语言，使数据加密时仍然能使用云计算环境，用户可在不需首次解密的情况下操纵加密的数据，从而使得对手拦截信息更加困难。

（8）视频、图像的检索和分析工具计划。旨在开发一个系统，能够利用军事图像分析员收集的数据进行大规模的军事图像分析。该项目如果成功，将使分析师能够在相关活动发生时建立警报。

此外，美国国防部高级研究计划局将开展XDATA项目的研究工作，计划在未来四年每年投入2500万美元开发能够分析海量半结构化数据（如表格数据、关系数据、分类数据、元数据）和非结构化数据（如文本文档、消息流量等）的计算技术和软件工具。需要解决的核心问题包括：开发可扩展的算法，用于处理分布式数据存储库中的不规则数据；创建有效的人机交互工具，用于支持面向各种处理任务的快速可定制视觉分析。XDATA项目将支持开放源软件工具集，以帮助开发机构灵活开发软件，使用户能够尽快实现海量数据处理能力，与特定国防应用的任务数据流保持同步。

3）美国国家卫生研究院——千人基因组计划的数据在亚马逊云上免费开放

美国国家卫生研究院宣布，由国际千人基因组计划创建的最大的人类遗传变异研究数据集在亚马逊网站云服务（AWS）上免费公开。截至目前，数据量已经达到大约200TB，相当于1600万个塞满文本资料的文件柜或3万多个标准DVD的容量。该数据集的规模极为庞大，几乎没有哪个研究机构具有足够的计算能力对其进行有效利用。现在AWS将千人基因组计划数据集免费公开，供研究人员自由访问和使用，他们只需为自己使用的计算服务付费。

4）美国能源部——通过高级计算技术加速科学发现

美国能源部将斥资2500万美元建立可扩展数据管理、分析与可视化（SDAV）研究所。在能源部劳伦斯伯克利国家实验室（Lawrence Berkeley National Laboratory）的组织下，SDAV研究所将汇集6大国家实验室和7所著名大学的专业知识和经验来开发新工具，帮助科学家对能源部超级计算机上的数据进行有效管理和可视化处理。此举将进一步简化和加速开发流程，使科学家们能够利用能源部的研究设施开展更加卓有成效的科学研究和发现。目前能源部的超级计算机上同时运行的数据流在规模和复杂性方面均呈现不断增长的势头，因此对这些新型工具的需求也更加迫切。开展的主要项目计划包括：

（1）高级科学计算研究办公室提供数据管理、可视化和数据分析集群，包括数字化保存和集群访问。

（2）高性能存储系统是对磁盘和磁带系统上PB级数据进行管理的数据管理软件。由美国能源部和IBM开发的高性能存储系统已在世界各地的大学和实验室投入使用，可用在数字图书馆以及宇航局和国会图书馆等部门。

（3）高性能存储系统能够对千万亿次的数据分析处理，从庞大的科学数据集提取信息，发现其主要特征并理解其间的关系。系统应用极为广泛，包括宇宙学和天气数据、传感器数据等。

（4）下一代网络方案支持工具使得在进行重大发现时能够实现合作研究。现今，每月有超过1PB的科学数据为开放式科学网格、地球系统网格等提供服务。工具中的中间件被得克萨斯大学、软件公司、石油公司使用，并一起培养学生利用先进的石油工程方法和集成的工作流程。

（5）基础能源科学办公室的科学用户设施旨在协助用户实现数据管理和分析大数据，可每天进行将一个单一的实验数据扩容到最大1012字节的数据的工作，可以最大限度地提高数据的可用性及更有效地利用同步加速器光源。



（6）生物和环境研究计划。大气辐射测量气候研究设施是一个多平台的科学用户设施，提供重要的大气现象的精确观测研究，主要用于应对从数以百计的文件中迅速采集和提交解决方法的挑战，以满足用户的需求。

（7）系统生物学知识库是一个社区驱动的软件框架，可实现对微生物、植物和环境条件下的生物群落功能的数据驱动的预测。系统生物学知识库是开放式的设计与开发环境，可以提高算法的开发和部署效率，并增加异构数据源的实验数据的获取和集成。

（8）通过聚变能源科学办公室和高级科学计算研究办公室合作，开发的数据管理技术，包括高性能的输入/输出系统、先进的科学的工作流程和框架、可视化技术的融合需求，已经吸引了欧洲一体化建模的注意，以及一个国际核聚变研究和工程项目的关注。

(9) 高能物理计算计划能够协助企业进行大数据管理,包括分布式分析产品以及高性能、快速、可扩展访问多种数据存储库的容错软件。

(10) 美国核数据计划是一个多方面努力,涉及 7 个国家实验室和两所大学的项目,提供跨越多个领域的专用数据库,包括核物理、编译和交叉检查以及对所有原子核的重要性质的相关实验结果等。

5) 美国地质勘探局——地球科学领域的大数据

约翰·威斯利·鲍威尔数据分析和合成中心致力于推动地球科学领域的思维创新和技术进步,它为科学家们创造了开展深层次分析、探索尖端计算功能和协作工具的机会和条件,这些功能和工具对于挖掘巨型数据集的价值具有重大意义。这些大数据项目将增进我们对许多重要问题的认识,包括气候变化、地震发生率和未来的生态指标对于地球物种产生的影响。

6) 国土安全部——卓越研究中心和可视化数据分析

通过对大量的异构数据进行研究,使得急救员可以发现人为或自然灾害、恐怖事件和需要执法的边境安全问题以及网络威胁的炸药。



7) 退伍军人管理部

由医疗保健信息研究所开发的自然语言处理工具,能够对在退伍军人管理部以文本形式存储的大量数据进行信息解锁。

目前,退伍军人管理部正在努力通过保护作战人员使用文字处理算法捕获公共卫生事件,正在开发一个的生产透明、可重复使用的各种安全相关的事件监控软件,以研究为基础的监控程序,能够跟踪、测量与军事部署相关的健康条件。

8) 卫生和人类服务部

(1) 疾病控制和预防中心。生物传感 2.0 是第一个在考虑区域和国家协调的情况下,通过互操作的网络系统对公众健康意识进行可行性分析的系统,它建立在现有的国家和地方的能力之上。生物传感 2.0 移除许多单片物理结构相关的成本,可对最终用户透明的分布式系统方面,以及做出适当的分析和报告的数据访问。

（2）医疗保险和医疗补助服务中心。基于 Hadoop 的一个数据仓库正处于研发阶段，它将支持医疗保险和医疗补助要求的分析和报告。其主要目标是建立可持续的、可扩展的设计，可容纳在数据仓库中进行积累，并补充现有的技术。

（3）食品与药品管理局。虚拟实验室环境将结合现有的资源和能力，使虚拟实验室数据网络具备先进的分析和统计工具，能够分析、预测和促进公众健康的文档管理，并且在世界范围内的合作中赋予亲临现场的能力，使任何地点在一小时内就具备在同一个虚拟实验室工作的功能。

9）国家档案和记录管理部

为十亿电子记录的网络基础设施提供一个联合机构主办的测试平台。这个多机构主办的网络基础设施将用于国家档案馆 87 万多样化的数字记录的文件和信息的收集，可称为计算研究所的“文艺复兴”。这个试验平台具有评估技术、方法和超大规模数据收集的功能，以支持可持续的访问。

10）国家航空和航天局



美国国家航空航天局的地球科学数据和信息系统项目将持续超过 15 年。它将致力于关注用户满意度，努力确保科学家和公众对数据的访问，以便从太空对地球进行研究，推动地球系统科学的发展，以应对气候和环境变化的挑战。

全球地球观测系统通过国际间努力合作进行地球观测数据共享和整合。美国宇航局已经与美国环境保护署、美国国家海洋、大气管理局以及其他机构和国家的力量强强联手，整合卫星、地面监测和建模系统，评估环境条件和预测人为和自然的森林火灾、人口增长和其他方面的结果。研究人员将在短期内整合各种复杂的空气质量信息，从而更好地了解和解决空气质量对环境和人体健康的影响。

太空行动协议由美国宇航局和 Cray 公司订立，在“大数据”系统的发展和低延迟的应用为中心的一个或多个项目上进行合作。该项目测试的混合计算机系统的实用程序能够使用一个高度集成的非 SQL 数据库作为数据传

输的手段，加速执行建模和分析软件。

太空望远镜科学研究所作为美国宇航局的分布式空间科学数据服务的一个组成部分，提供多任务存档的支持，及各种天文数据档案和相关的科学数据，包括光学、紫外线、近红外光谱等光学相关的科学数据。太空望远镜科学研究所支持多种工具，可以对各种光谱图像数据进行访问。

11）国家人文基金会

数据挖掘的挑战旨在分析大数据的变化对人文社会科学的影响，这种新的计算为基础的研究方法需要搜索、分析和理解大量的材料，如数字化的书籍和报纸数据库，利用网络搜索、传感器和手机记录交易数据。在国家人文基金会的领导下，这个挑战性的研究工作将由美国和 8 个国际组织在四个国家中进行。

12）国家安全局

竞争网络防御规模的情境意识的培养和测试将探索数据可视化的网上竞赛，从开始与识别此类事件设计与最佳执行方法上，进行庞大的计算机网络防御上的数据可视化开发。



情报共同体通过与美国政府、学术界和工业界各种各样的合作伙伴确定一套协调、宣传和活动方案，将网络安全与大数据相结合，使学术界了解其观点。

国家安全局/中央安全服务部的商业解决方案中心通过供应商的能力演示，展示了新的商业技术的发展，以满足国家安全局/ 中央安全服务部和国家安全社区的战略需求。

5. 中国政府大数据战略的主要项目介绍

1）政府数据资源共享开放工程

其目的是：推动政府数据资源共享。制定政府数据资源共享管理办法。

形成政府数据统一共享交换平台。到 2018 年，在中央政府层面实现金税、金关、金财、金审、金盾、金宏、金保、金土、金农、金水、金质等信息系统通

过统一平台进行数据共享和交换。

形成国家政府数据统一开放平台。建立政府部门和事业单位等公共机构数据资源清单，制定实施政府数据开放共享标准，制订数据开放计划。

2）国家大数据资源统筹发展工程

整合各类政府信息平台和信息系统。在地市级以上（含地市级）政府集中构建统一的互联网政务数据服务平台和信息惠民服务平台。

整合分散的数据中心资源。构建形成布局合理、规模适度、保障有力、绿色集约的政务数据中心体系。开展区域试点。

加快完善国家基础信息资源体系。到2018年，跨部门共享校核的国家人口基础信息库、法人单位信息资源库、自然资源和空间地理基础信息库等国家基础信息资源体系基本建成。

加强互联网信息采集利用。制订、完善互联网信息保存相关法律法规，构建互联网信息保存和信息服务体系。

3）政府治理大数据工程



推动宏观调控决策支持、风险预警和执行监督大数据应用。探索建立国家宏观调控决策支持、风险预警和执行监督大数据应用体系。

推动信用信息共享机制和信用信息系统建设。鼓励互联网企业运用大数据技术建立市场化的第三方信用信息共享平台，建设企业信用信息公示系统，初步建成社会信用体系。

建设社会治理大数据应用体系。实时采集并汇总分析政府部门和企事业单位的市场监管、检验检测、违法失信、企业生产经营、销售物流、投诉举报、消费维权等数据。

4）公共服务大数据工程

医疗健康服务大数据。建设覆盖公共卫生、医疗服务、医疗保障、药品供应、计划生育和综合管理业务的医疗健康管理和服务大数据应用体系。

社会保障服务大数据。建设由城市延伸到农村的统一社会救助、社会福利、社会保障大数据平台。

教育文化大数据。建立各阶段适龄入学人口基础数据库、学生基础数据库和终身电子学籍档案。

交通旅游服务大数据。建立综合交通服务大数据平台。建立旅游投诉及评价全媒体交互中心。

5）工业和新兴产业大数据工程

工业大数据应用。研究推动大数据在研发设计、生产制造、经营管理、市场营销、售后服务等产业链各环节的应用。

服务业大数据应用。研发面向服务业的大数据解决方案。

培育数据应用新业态。大力培育互联网金融、数据服务、数据处理分析、数据影视、数据探矿、数据化学、数据材料、数据制药等新业态。

电子商务大数据应用。电子商务企业应依法向政府部门报送数据。

6）现代农业大数据工程

农业农村信息综合服务。建设农产品全球生产、消费、库存、进出口、价格、成本等数据调查分析系统工程，构建面向农业农村的综合信息服务平台。



农业资源要素数据共享。建立我国农业耕地、草原、林地、水利设施、水资源、农业设施设备、新型经营主体、农业劳动力、金融资本等资源要素数据监测体系。

农产品质量安全信息服务。建立农产品生产的生态环境、生产资料、生产过程、市场流通、加工储藏、检验检测等数据共享机制。

7）万众创新大数据工程

大数据创新应用。鼓励企业和公众发掘利用开放数据资源。

大数据创新服务。研发一批大数据公共服务产品。

发展科学大数据。构建科学大数据国家重大基础设施。发展科学大数据应用服务中心。

知识服务大数据应用。建立国家知识服务平台与知识资源服务中心。

8）大数据关键技术及产品研发与产业化工程

加强大数据基础研究。探讨建立数据科学的学科体系;研究面向大数据计算的新体系和大数据分析理论,探索建立数据科学驱动行业应用的模型。

大数据技术产品研发。加强数据存储、整理、分析处理、可视化、信息安全与隐私保护等领域技术产品的研发。

提升大数据技术服务能力。以应用带动大数据技术和产品研发,形成面向各行业的成熟的大数据解决方案。

9）大数据产业支撑能力提升工程

培育骨干企业。到 2020 年,培育 10 家国际领先的大数据核心龙头企业,500 家大数据应用、服务和产品制造企业。

大数据产业公共服务。形成面向大数据相关领域的公共服务平台。

中小微企业公共服务大数据。形成全国统一的中小微企业公共服务大数据平台。

10）网络和大数据安全保障工程

网络和大数据安全支撑体系建设。到 2020 年,实现关键部门的关键设备安全可靠。完善网络安全保密防护体系。

大数据安全保障体系建设。建设完善金融、能源、交通、电信、统计、广电、公共安全、公共事业等重要数据资源和信息系统的安全保密防护体系。

网络安全信息共享和重大风险识别大数据支撑体系建设。建立网络安全信息共享机制,推动政府、行业、企业间的网络风险信息共享。



附录B 《G8开放数据宪章》及其技术附件要点

B.1 《G8开放数据宪章》要点

1. 开放数据是这场全球性运动的核心所在

开放数据是一个具有巨大潜力的未开发资源。它有助于建设一个更加强大、更加相互关联、更好满足公民需求、激励创新和蓬勃发展的社会。

获取、发布和再利用G8政府提供的数据的基础原则是：**默认开放数据；注重质量和数量；让所有人都可用；为改善治理发布数据；为激发创新发布数据**。

开放数据的益处能够而且应该为各国公民享有。



2. 原则

1）原则1：默认开放数据

开放数据的免费获取以及后续的再利用有着重要的社会和经济价值。政府要向默认开放数据的方向转变。

“政府数据”这一术语含义广泛，可以适用于国家、联邦、地方、国际政府机构或更广泛的公共部门拥有的数据。

必须遵守各国和国际的法律法规，尤其是关于知识产权、个人身份和敏感信息的法律法规。

希望所有政府数据能以默认方式公开发布。同时我们也认识到，由于正当原因，一些数据不能发布。

2）原则2：注重质量和数量

政府和公共部门持有的大量信息可能是公民感兴趣的。

准备高质量的数据可能需要时间，并且与各方、国家间以及更广泛的开

放数据用户进行协商确定哪些数据优先发布或改进是重要的。

- 发布及时、全面、准确的高质量开放数据。尽可能地使数据保持其原始的、未经改动的形式和最好的颗粒度。
- 确保数据中的信息以简单、清晰的语言描述，使其可以被所有人理解，但本宪章并不要求翻译成其他语言。
- 确保数据都被充分说明，让消费者有足够的信息来了解数据的优势、劣势、分析的局限性和安全要求，以及如何处理数据。
- 尽早发布数据，允许用户提供反馈，然后持续进行修订，确保开放数据质量满足最高标准。

3）原则 3：让所有人都可用

数据的发布方式应有助于所有人能够获取和再利用数据。

开放数据应该是免费提供的，以鼓励它们被最广泛地使用。

发布开放数据时，应当没有诸如注册登记等阻止人们访问数据的官僚或行政障碍。

- 尽可能以开放格式发布数据，确保数据被最广泛的用户在最广泛的用途中使用；
- 尽可能多地发布数据，对于现在不能免费提供的数据，增加补贴，以鼓励其免费提供。在许多情况下，这将包括提供多种格式的数据，以便它们可以用计算机处理并被人们所理解。

4）原则 4：为改善治理发布数据

开放数据的发布有利于加强民主制度建设和促进更好的政策制定，以满足公民需求。这不仅在我们自己的国家如此，在世界各地都是如此。

其他多边组织和机构对开放数据的兴趣越来越大。

- 与各方及世界其他国家分享技术和经验，让每个人都能从开放数据中获益。
- 通过在线记录所有相关流程，确保数据采集、标准和发布过程的透明。



5）原则 5：为激励创新发布数据

认识到多样性在激励创造力和创新方面的重要性，我们同意，使用我们的数据的个人和组织越多，产生的社会和经济效益就越大。这对于商业和非商业用途都是适用的。

- 努力营造开放数据文化并鼓励应用程序开发者、从事开放数据推广工作的民间社会组织等挖掘开放数据的价值。
- 通过以机器可读的格式提供数据，壮大未来数据创新者队伍。

B.2 《G8 开放数据宪章》技术附件要点

1. 最佳实践

1）原则 1：默认开放数据

认识到开放数据的重要性，希望所有政府数据能以默认方式公开发布。



- 在公开的声明如公告、战略或政策中明确开放数据的定位，以使推进开放数据的计划进程在司法上是清晰明确的。
- 发布一个国家行动计划，依据《G8 开放数据宪章》的原则，细化开放数据的计划。
- 在国家的门户网站上发布数据，使所有已经公布的政府数据可以很容易在一个地方找到。门户可能是一个可以从上面下载数据的中心网站，或是一个列出所有存储在不同位置的政府开放数据的网站。每个门户网站将包括一个列出所有数据和元数据的注册表文件，同时为开发者提供应用编程接口（APIs）。如果不可能在一个门户网站上公布所有数据，那就要清楚地展现数据位置，而且在没有通知的情况下不能进行转移。

2）原则 2：注重质量和数量

发布的数据既要注重质量，又要注重数量。以有助于人们理解并使用的方式发布数据，这将有助于提高不同政策领域、企业或国家数据的互操

作性。

- 使用稳定和一致的元数据(即描述实际数据的字段或元素)。
- 在G8内发布和维护一个最新的核心描述性元数据字段映射,让来自世界各地的人们能够更容易使用和理解。这将使目前还没有一个数据门户的G8或非G8国家,考虑采用这个映射中包含的元数据字段。
- 确保数据描述充分,以帮助用户充分了解数据。这可以包括:提供数据字段使用说明的文档;链接不同数据的数据字典;一个描述数据采集目的、目标受众、样品特征,以及数据采集方法的用户指南。
- 倾听数据用户的反馈,以改善所提供数据的广度、质量和可访问性。这可通过国家数据战略或政策的公共咨询、与民间组织的讨论、在数据门户网站建立反馈机制等形式,或通过其他的适当机制来实现。

3)原则3:让所有人都可用

数据的发布方式应有助于所有人能够获取和再利用数据。

- 以方便的开放格式提供数据,确保文件可以很容易地被所有常用的网络搜索应用工具检索、下载、索引和查找。开放的格式是指那些可供任何人免费使用的规范格式,如非专有的纯文本逗号分隔符文件(CSV),从而使文件中包含的数据能够被不同的软件程序打开。

4)原则4:为改善治理发布数据

数据是可以提升政府效能、效率和快速响应公民需求的有力工具,同时又能进一步激发开放数据的需求。

- 与民间社会组织和个人建立联系,让公众反馈他们最想要政府发布的数据。
- 为了使数据标准更加开放,应该考虑:其他国家和国际组织发布的数据;来自其他国际增加透明度行动的标准。
- 记录我们在开放数据工作上的经验,例如,发布开放数据政策、实践和门户网站的技术信息,使其他国家共享开放数据的益处。



5）原则 5：为激发创新发布数据

公民可以在自己的国家和世界范围内使用我们的数据来推动创新。免费获取和再利用开放政府数据是推动创新的基本因素。

- 支持使用开放许可证或者其他相关措施发布数据——同时尊重知识产权——这样除特殊情况下，针对商业和非商业目的的信息再利用将不受限制或者免费。
- 提供结构良好的数据以确保机器可以批量读取，从而使自动处理和访问时需要下载的文件最少。
- 使用应用程序接口（APIs）发布数据，并在适当的情况下，确保经常定期更新和访问的数据可以被便捷地获取。
- 通过各国组织竞赛、奖励或者指导数据用户等形式，鼓励创新使用我们的数据。

2. 共同行动

1）行动 1：G8 国家行动计划

发布各国的国家行动计划，详细介绍各国如何依据自己的国家框架执行《G8 开放数据宪章》（2013 年 10 月）。

报告年度进展（通过 G8 问责工作组）（2014 年和 2015 年）。

2）行动 2：发布高价值数据

表 B.1 中的领域的数据对改善民主和鼓励创新性的数据再利用具有很高的价值。

表 B.1　数据分类与数据集示例

数据分类（按字母顺序排列）	数据集示例
公司	公司/企业登记
犯罪与司法	犯罪统计、安全
地球观测	气象/天气、农业、林业、渔业和狩猎
教育	学校名单、学校表现、数字技能
能源与环境	污染程度、能源消耗

续表

数据分类(按字母顺序排列)	数据集示例
财政与合同	交易费用、合约、招标、邀标、地方预算、国家预算(计划和支出)
地理空间	地形、邮政编码、国家地图、本地地图
全球发展	援助、粮食安全、采掘业、土地
政府问责与民主	政府联络点、选举结果、法律法规、薪金(薪级)、招待/礼品
健康	处方数据、效果数据
科学与研究	基因组数据、研究和教育活动、实验结果
统计	国家统计、人口普查、基础设施、财产、从业人员
社会流动性与福利	住房、医疗保险和失业救济
交通运输与基础设施	公共交通时间表、宽带接入点及普及率

按照“默认开放数据”和“注重质量和数量”的原则,积极推进这些数据的开放。

第一步将共同推进有关国家统计、国家地图、国家选举和国家预算的关键数据集的建设和发布(自 2013 年 6 月起),同时,努力改善其颗粒度和可访问性(截至 2013 年 12 月)。



所有 G8 成员的共同行动有助于消除障碍和提出创新的解决方案,以应对面临的挑战。各方要共同努力,加大国家关键领域(如民主和环境[①]等)的政府开放数据的提供力度。

根据各国的国家框架,在各国的国家行动计划中对如何以及何时发布其余类别数据进行阐述(2013 年 10 月)。

3) 行动 3:元数据映射

继续维护 G8 元数据映射的实践活动(2013 年 6 月)。该映射可以在 Github 网站查看,包括一个横跨 G8 成员的元数据映射索引集合和一个有关各 G8 成员在其国家门户使用元数据的详细页面。

① 目录和数据集最终确定于 2013 年 12 月。

附录 C 信息协同服务接口的 XML Schema 描述

1. 数据库协同请求对象 XML Schema 描述

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
elementFormDefault="qualified">
    <xs:element name="DBExRequest">
        <xs:complexType>
            <xs:sequence>
                <xs:element ref="version"/>
                <xs:element ref="senderCode"/>
                <xs:element ref="senderAppCode"/>
                <xs:element ref="resourceID"/>
                <xs:element ref="resourceName"/>
                <xs:element ref="userID"/>
                <xs:element ref="userName"/>
                <xs:element ref="timeStamp"/>
                <xs:element ref="messageID"/>
                <xs:element ref="receiver" maxOccurs="unbounded"/>
            </xs:sequence>
        </xs:complexType>
    </xs:element>
    <xs:element name="messageID" type="xs:string"/>
    <xs:element name="receiver" type="ReceiverType"/>
    <xs:element name="receiverAppCode" type="xs:string"/>
    <xs:element name="receiverCode" type="xs:string"/>
    <xs:element name="resourceID" type="xs:string"/>
    <xs:element name="resourceName" type="xs:string"/>
    <xs:element name="senderAppCode" type="xs:string"/>
    <xs:element name="senderCode" type="xs:string"/>
    <xs:element name="timeStamp" type="xs:string"/>
    <xs:element name="userID" type="xs:string"/>
    <xs:element name="userName" type="xs:string"/>
    <xs:element name="version" type="xs:string"/>
```

```
    <xs:complexType name="ReceiverType">
        <xs:sequence>
            <xs:element ref="receiverCode"/>
            <xs:element ref="receiverAppCode"/>
        </xs:sequence>
    </xs:complexType>
</xs:schema>
```

数据库协同请求对象参数见表 C.1。

表 C.1 数据库协同请求对象参数

参数名称	参数类型	参数说明
version	String	服务版本号
senderCode	String	发送方机构编码
senderAppCode	String	发送方应用系统编码
resourceID	String	资源标识符
resourceName	String	资源名称
userID	String	用户标识符
userName	String	用户名称
timeStamp	String	时间戳
messageID	String	消息标识符
receiver	ReceiverType	接收方

接收方参数见表 C.2。

表 C.2 接收方(ReceiverType)参数

参数名称	参数类型	参数说明
receiverCode	String	接收方机构编码
receiverAppCode	String	接收方应用系统编码

2. 数据库协同数据对象 XML Schema 描述

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
elementFormDefault="qualified">
```




```
<xs:element name="DBExData">
    <xs:complexType>
        <xs:sequence>
            <xs:element ref="senderAppCode"/>
            <xs:element ref="senderCode"/>
            <xs:element ref="senderPeerName"/>
            <xs:element ref="receiverAppCode"/>
            <xs:element ref="receiverCode"/>
            <xs:element ref="receiverPeerName"/>
            <xs:element ref="sendTime"/>
            <xs:element ref="resourceID"/>
            <xs:element ref="resourceType"/>
            <xs:element ref="userName"/>
            <xs:element ref="userID"/>
            <xs:element ref="processName"/>
            <xs:element ref="dataSet"/>
        </xs:sequence>
    </xs:complexType>
</xs:element>
<xs:element name="senderAppCode" type="xs:string"/>
<xs:element name="senderCode" type="xs:string"/>
<xs:element name="senderPeerName" type="xs:string"/>
<xs:element name="sendTime" type="xs:string"/>
<xs:element name="userID" type="xs:string"/>
<xs:element name="userName" type="xs:string"/>
<xs:element name="operationType" type="xs:string"/>
<xs:element name="processName" type="xs:string"/>
<xs:element name="receiverAppCode" type="xs:string"/>
<xs:element name="receiverCode" type="xs:string"/>
<xs:element name="receiverPeerName" type="xs:string"/>
<xs:element name="resourceID" type="xs:string"/>
<xs:element name="resourceType" type="xs:string"/>
<xs:element name="dataSet" type="DataSetType"/>
<xs:element name="recordData" type="RecordDataType"/>
<xs:element name="unitData" type="UnitDataType"/>
<xs:complexType name="DataSetType">
    <xs:sequence>
        <xs:element name="operationType" type="xs:string"/>
        <xs:element ref="recordData" maxOccurs="unbounded"/>
```

```
        </xs:sequence>
    </xs:complexType>
    <xs:complexType name="RecordDataType">
        <xs:sequence>
            <xs:element ref="unitData" maxOccurs="unbounded"/>
        </xs:sequence>
    </xs:complexType>
    <xs:complexType name="UnitDataType">
        <xs:sequence>
            <xs:element name="unitIDName" type="xs:string"/>
            <xs:element name="unitDisplayName" type="xs:string"/>
            <xs:element name="unitValue" type="xs:string"/>
        </xs:sequence>
    </xs:complexType>
</xs:schema>
```

数据库协同数据对象参数见表 C. 3。

表 C. 3　数据库协同数据对象参数

参 数 名 称	参数类型	参 数 说 明
senderAppCode	String	发送方应用系统编码
senderCode	String	发送方机构编码
senderPeerName	String	发送方节点名称
receiverAppCode	String	接收方应用系统编码
receiverCode	String	接收方机构编码
receiverPeerName	String	接收方节点名称
sendTime	String	发送时间
resourceID	String	资源标识符
resourceType	String	资源类型
userID	String	用户标识符
userName	String	用户名
processName	String	交换流程名称
dataSet	DataSetType	交换的数据集





数据集、记录、数据项参数分别见表 C.4、表 C.5 和表 C.6。

表 C.4 数据集(DataSetType)参数

参数名称	参数类型	参数说明
operationType	String	操作码,I—增加 U—修改 D—删除
recordData	RecordDataType	组成数据集的基本单位,表示一条记录

表 C.5 记录(RecordDataType)参数

参数名称	参数类型	参数说明
unitData	UnitDataType	数据项,组成数据记录的基本单位,表示关系数据库表中的某个字段

表 C.6 数据项(UnitDataType)参数

参数名称	参数类型	参数说明
unitIDName	String	数据项的标识符
unitDisplayName	String	数据项的名称
unitValue	String	数据项的值

3. 文件协同请求对象 XML Schema 描述

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
elementFormDefault="qualified">
    <xs:element name="FileExRequest">
        <xs:complexType>
            <xs:sequence>
                <xs:element ref="version"/>
                <xs:element ref="senderCode"/>
                <xs:element ref="senderAppCode"/>
                <xs:element ref="resourceID"/>
                <xs:element ref="resourceName"/>
                <xs:element ref="userID"/>
                <xs:element ref="userName"/>
                <xs:element ref="timeStamp"/>
```

```
                <xs:element ref="messageID"/>
                <xs:element ref="sourceFile" maxOccurs="unbounded"/>
                <xs:element ref="receiver" maxOccurs="unbounded"/>
            </xs:sequence>
        </xs:complexType>
    </xs:element>
    <xs:element name="messageID" type="xs:string"/>
    <xs:element name="receiver" type="ReceiverType"/>
    <xs:element name="receiverAppCode" type="xs:string"/>
    <xs:element name="receiverCode" type="xs:string"/>
    <xs:element name="resourceID" type="xs:string"/>
    <xs:element name="resourceName" type="xs:string"/>
    <xs:element name="senderAppCode" type="xs:string"/>
    <xs:element name="senderCode" type="xs:string"/>
    <xs:element name="sourceFile" type="SourceFileType"/>
    <xs:element name="sourceFileName" type="xs:string"/>
    <xs:element name="sourceFilePath" type="xs:string"/>
    <xs:element name="timeStamp" type="xs:string"/>
    <xs:element name="userID" type="xs:string"/>
    <xs:element name="userName" type="xs:string"/>
    <xs:element name="version" type="xs:string"/>
    <xs:complexType name="ReceiverType">
        <xs:sequence>
            <xs:element ref="receiverCode"/>
            <xs:element ref="receiverAppCode"/>
        </xs:sequence>
    </xs:complexType>
    <xs:complexType name="SourceFileType">
        <xs:sequence>
            <xs:element ref="sourceFileName"/>
            <xs:element ref="sourceFilePath"/>
        </xs:sequence>
    </xs:complexType>
</xs:schema>
```

文件协同请求对象参数见表 C.7。

表 C.7　文件协同请求对象参数

参数名称	参数类型	参数说明
version	String	服务版本号
senderCode	String	发送方机构编码
senderAppCode	String	发送方应用系统编码
resourceID	String	资源标识符
resourceName	String	资源名称
userID	String	用户标识符
userName	String	用户名称
timeStamp	String	时间戳
messageID	String	消息标识符
sourceFile	SourceFileType	交换的文件数据
receiver	ReceiverType	接收方

源文件及接收方参数分别见表 C.8 和表 C.9。

表 C.8　源文件(SourceFileType)参数

参数名称	参数类型	参数说明
sourceFileName	String	文件名
sourceFilePath	String	文件路径

表 C.9　接收方(ReceiverType)参数

参数名称	参数类型	参数说明
receiverCode	String	接收方机构编码
receiverAppCode	String	接收方应用系统编码

4. 文件协同数据对象 XML Schema 描述

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
elementFormDefault="qualified">
    <xs:element name="FileExData">
```

```
        <xs:complexType>
            <xs:sequence>
                <xs:element ref="senderAppCode"/>
                <xs:element ref="senderCode"/>
                <xs:element ref="senderPeerName"/>
                <xs:element ref="receiverAppCode"/>
                <xs:element ref="receiverCode"/>
                <xs:element ref="receiverPeerName"/>
                <xs:element ref="sendTime"/>
                <xs:element ref="resourceID"/>
                <xs:element ref="resourceType"/>
                <xs:element ref="userName"/>
                <xs:element ref="userID"/>
                <xs:element ref="processName"/>
                <xs:element ref="dataSet"/>
            </xs:sequence>
        </xs:complexType>
    </xs:element>
    <xs:element name="dataSet" type="DataSetType"/>
    <xs:element name="isBinary" type="xs:boolean"/>
    <xs:element name="fileName" type="xs:string"/>
    <xs:element name="processName" type="xs:string"/>
    <xs:element name="receiverAppCode" type="xs:string"/>
    <xs:element name="receiverCode" type="xs:string"/>
    <xs:element name="receiverPeerName" type="xs:string"/>
    <xs:element name="resourceID" type="xs:string"/>
    <xs:element name="resourceType" type="xs:string"/>
    <xs:element name="senderAppCode" type="xs:string"/>
    <xs:element name="senderCode" type="xs:string"/>
    <xs:element name="senderPeerName" type="xs:string"/>
    <xs:element name="sendTime" type="xs:string"/>
    <xs:element name="fileSize" type="xs:string"/>
    <xs:element name="fileType" type="xs:string"/>
    <xs:element name="userID" type="xs:string"/>
    <xs:element name="userName" type="xs:string"/>
    <xs:element name="unitData" type="xs:string"/>
    <xs:complexType name="DataSetType">
```



```
        <xs:sequence>
    <xs:element name="recordData" type="RecordDataType"
    maxOccurs="unbounded"/>
        </xs:sequence>
    </xs:complexType>
    <xs:complexType name="RecordDataType">
        <xs:sequence>
            <xs:element ref="fileName"/>
            <xs:element ref="fileType"/>
            <xs:element ref="isBinary"/>
            <xs:element ref="fileSize"/>
            <xs:element ref="unitData"/>
        </xs:sequence>
    </xs:complexType>
</xs:schema>
```

文件协同数据对象参数见表 C.10。

表 C.10　文件协同数据对象参数

参数名称	参数类型	参数说明
senderAppCode	String	发送方应用系统编码
senderCode	String	发送方机构编码
senderPeerName	String	发送方节点名称
receiverAppCode	String	接收方应用系统编码
receiverCode	String	接收方机构编码
receiverPeerName	String	接收方节点名称
sendTime	String	发送时间
resourceID	String	资源标识符
resourceType	String	资源类型
userName	String	用户名称
userID	String	用户标识符
processName	String	流程名称
dataSet	DataSetType	交换的文件数据集

数据集参数见表 C.11。

表 C.11　数据集(DataSetType)参数

参数名称	参数类型	参数说明
recordData	RecordDataType	组成数据集的基本单位,表示一个文件

记录参数见表 C.12。

表 C.12　记录(RecordDataType)参数

参数名称	参数类型	参数说明
fileName	String	文件名称
fileType	String	文件类型(doc、xls 等)
isBinary	Boolean	是否为二进制文件,TRUE 表示二进制,FALSE 表示文本格式
fileSize	String	文件大小
unitData	String	经过 base64 编码后的文件内容



附录D 基于MATLAB的模糊聚类核心计算程序

1. 数据标准化变换

```
function[X]=F_1Y (cs, X)
    if(cs==0) return; end
    [n, m]=size(X);
    if(cs==1)        %平移-标准差变换
        for(k=1:m) xk=0;
            for(i=1:n) xk=xk+X(i, k); end
            xk=xk/n; sk=0;
            for(i=1:n) sk=sk+ (X(i, k)-xk) ^2; end
            sk=sqrt(sk/n);
            for(i=1:n) X(i, k)= (X(i, k)-xk)/sk; end; end
    else           %平移-极差变换
        for (k=1:m)
           xmin=X(1, k); xmax=X(1, k);
           for(i=1:n)
               if(xmin>X(i, k)) xmin=X(i, k); end
               if(xmax<X(i, k)) xmax=X(i, k); end; end
           for(i=1:n) X(i, k)= (X(i, k)-xmin)/(xmax-xmin); end; end; end
```

2. 建立模糊相似矩阵 *R*

```
function[R]=F_2R(cs, X)
[n, m]=size(X); R=[];
if(cs==1)        %数量积法
    maxM=0; pd=0;
    for (i=1:n) for(j=1:n)
            if(j~=i) x=0;
                for(k=1:m) x=x+X(i, k) * X(j, k); end
                if(maxM<x) maxM=x; end; end; end; end
    if (maxM<0.000001) return; end
    maxM=maxM+1;
```

```
    for(i=1:n) for(j=1:n)
            if(i==j) R(i, j)=1;
            else
                R(i, j)=0;
                for(k=1:m) R(i, j)=R(i, j)+X(i, k) * X(j, k); end
                R(i, j)=R(i, j)/maxM;
                if(R(i, j)<0) pd=1; end; end; end; end
    if (pd) for(i=1:n) for(j=1:n) R(i, j)= (R(i, j)+1)/2; end; end; end
elseif(cs==2)       %夹角余弦法
    for(i=1:n) for(j=1:n) xi=0; xj=0;
            for(k=1:m) xi=xi+X(i, k) ^ 2; xj=xj+X(j, k) ^ 2; end
            s=sqrt(xi * xj);
            R(i, j)=0;
            for(k=1:m) R(i, j)=R(i, j)+X(i, k) * X(j, k); end
            R(i, j)=R(i, j)/s; end; end
elseif(cs==3)       %相关系数法
    for(i=1:n) for(j=1:n) xi=0; xj=0;
            for(k=1:m) xi=xi+X(i, k); xj=xj+X(j, k); end
            xi=xi/m; xj=xj/m; xis=0; xjs=0;
            for(k=1:m) xis=xis+ (X(i, k)-xi)  ^ 2; xjs=xjs+ (X(j, k)-xj) ^ 2; end
            s=sqrt(xis * xjs);
            R(i, j)=0;
            for(k=1:m) R(i, j)=R(i, j)+abs((X(i, k)-xi) * (X(j, k)-xj)); end
            R(i, j)=R(i, j)/s; end; end
elseif(cs==4)       %指数相似系数法
    for(i=1:n) for(j=1:n) R(i, j)=0;
            for(k=1:m) xk=0;
                for(z=1:n) xk=xk+X(z, k); end
                xk=xk/n; sk=0;
                for(z=1:n) sk=sk+ (X(z, k)-xk) ^ 2; end
                sk=sk/n;
                R(i, j)=R(i, j)+exp(-0.75* ((X(i, k)-X(j, k))/sk) ^ 2); end
            R(i, j)=R(i, j)/m; end; end
elseif(cs<=7)       %最大最小值法、算术平均最小法、几何平均最小法
    for(i=1:n) for(j=1:n) fz=0; fm=0;
            for(k=1:m)
                if(X(j, k)<0) R=[]; return; end
                if(X(j, k)<X(i, k)) x=X(i, k); else x=X(j, k); end
                fz=fz+x; end
```

```
            if(cs==5)      %最大最小值法
                for(k=1:m) if(X(i, k)>X(j, k)) x=X(j, k); else x=X(j, k); end; end
                fm=fm+x;
            elseif(cs==6)      %算术平均最小法
                for(k=1:m) fm=fm+(x(i, k)+X(j, k))/2; end
            else          %几何平均最小法
                for(k=1:m) fm=fm+sqrt(X(i, k)*X(j, k)); end; end
            R(i, j)=fz/fm; end; end
elseif(cs<=10) C=0;     %直接距离法
    for(i=1:n) for(j=i+1:n) d=0;
            if(cs==8)      ; %欧几里得距离
                for(k=1:m) d=d+(X(i, k)-X(j, k))^2; end
                d=sqrt(d)
            elseif(cs==9)      %海明距离
                for(k=1:m) d=d+abs(X(i, k)-X(j, k)); end
            else    %切比雪夫距离
                for(k=1:m) if(d<abs(X(i, k)-X(j, k))) d=abs(X(i, k)-X(j, k)); end;
end; end
            if(C<d) C=d; end; end; end
    C=1/(1+C);
    for(i=1:n) for(j=1:n) d=0;
            if(cs==8)      %欧几里得距离
                for(k=1:m) d=d+(X(i, k)-X(j, k))^2; end
                d=sqrt(d);
            elseif(cs==9)      %海明距离
                for(k=1:m) d=d+abs(X(i, k)-X(j, k)); end
            else          %切比雪夫距离
                for(k=1:m) if(d<abs(X(i, k)-X(j, k))) d=abs(X(i, k)-X(j, k)); end;
end; end
            R(i, j)=1-C*d; end; end
elseif(cs<=13)     %倒数距离法
    minM=Inf;
    for(i=1:n) for(j=i+1:n) d=0;
            if(cs==11)      %欧几里得距离
                for(k=1:m) d=d+(X(i, k)-X(j, k))^2; end
                d=sqrt(d)
            elseif(cs==12)      %海明距离
                for(k=1:m) d=d+abs(X(i, k)-X(j, k))^2; end
            else          %切比雪夫距离
```

```
                    for(k=1:m) if(d<abs(X(i, k)-X(j, k))) d=abs(X(i, k)-X(j, k));
end; end; end
                if(minM>d) minM=d; end; end; end
    minM=0.9999*minM;
    if(minM<0.000001) return; end
    for(i=1:n) for(j=1:n) d=0;
                if(j==i) R(i, j)=1; continue; end
                if(cs==11)      %欧几里得距离
                    for(k=1:m) d=d+(X(i, k)-X(j, k))^2; end
                    d=sqrt(d);
                elseif(cs==12)      %海明距离
                    for(k=1:m) d=d+abs(X(i, k)-X(j, k)); end
                else            %切比雪夫距离
                    for(k=1:m)
                        if(d<abs(X(i, k)-X(j, k))) d=abs(X(i, k)-X(j, k)); end;
end; end
                R(i, j)=minM/d;
        end; end
else            %指数距离法
    for(i=1:n) for(j=1:n) d=0;
                if(cs==14)      %欧几里得距离
                    for(k=1:m) d=d+(X(i, k)-X(j, k))^2; end
                    d=sqrt(d);
                elseif(cs==15)      %海明距离
                    for(k=1:m) d=d+abs(X(i, k)-X(j, k)); end
                else            %切比雪夫距离
                    for(k=1:m) if(d<abs(X(i, k)-X(j, k))) d=abs(X(i, k)-X(j, k));
end; end; end
                R(i, j)=exp(-d); end; end; end
```

3. 矩阵的合成运算(先取大,后取小)

```
function[C]=Max_Min(A, B)
[m, s]=size(A);
[s1, n]=size(B);
C=[];
if (s1~=s) return; end
for (i=1:m) for(j=1:n)
    C(i, j)=0;
```

```
    For (k=1:s) x=0;
        if(A(i, k)<B(k, j)) x=A(i, k);
        else x=B(k, j); end
        if (C(i, j)<x) C(i, j)=x; end; end; end; end
```

4. 动态聚类

```
function F_3BD(R)
[m, n]=size(R);
if(m~=n | m==0) return; end
for(i=1:n) R(i, i)=1;        %修正错误
    for(j=i+1:n)
        if(R(i, j)<0) R(i, j)=0; elseif(R(i, j)>1) R(i, j)=1; end
        R(i, j)=round(10000*R(i, j))/10000; R(j, i)=R(i, j); end; end
js0=0;
while(1)             %求传递闭包
    R1=Max_Min(R, R); js0=js0+1
    if(R1==R) break; else R=R1; end; end
lmd(1)=1; k=1;
for(i=1:n) for(j=i+1:n) pd=1;        %找出所有不相同的元素
        for(x=1:k) if(R(i, j)==lmd(x)) pd=0; break; end; end
        if(pd) k=k+1; lmd(k)=R(i, j); end; end; end
for(i=1:k-1) for(j=i+1:k)        %从大到小排序
        if(lmd(i)<lmd(j)) x=lmd(j); lmd(j)=lmd(i); lmd(i)=x; end; end; end
for(x=1:k) js=0; flsz(x)=0;        %按 lmd(x) 分类,分类数为 flsz(x),临时用 Sz 记录
元素序号
    for(i=1:n)  pd=1;
        for(y=1:js)  if(Sz(y)==i) pd=0; break; end; end
        if(pd)
            for(j=1:n) if(R(i, j)>=lmd(x)) js=js+1; Sz(js)=j; end; end
            flsz(x)=flsz(x)+1; end; end; end
for(i=1:k-1) for(j=i+1:k)
        if(flsz(j)==flsz(i)) flsz(j)=0; end; end; end
fl=0;        %排除相同的分类
for(i=1:k) if(flsz(i)) fl=fl+1; lmd(fl)=lmd(i); end; end
for(i=1:n) xhsz(i)=i; end
for(x=1:fl) js=0; flsz(x)=0;        %获得分类情况:对分类元素进行排序
    for(i=1:n) pd=1;
        for(y=1:js) if(Sz(y)==i) pd=0; break; end; end
```

```
        if(pd)
            if(js==0) y=0; end
            for(j=1:n) if(R(i, j)>=lmd(x)) js=js+1; Sz(js)=j; end; end
            flsz(x)=flsz(x)+1; Sz0(flsz(x))=js-y; end; end
    js0=0;
    for(i=1:flsz(x))
        for(j=1:Sz0(i)) Sz1(j)=Sz(js0+j); end
        for(j=1:n) for(y=1:Sz0(i))
                if (xhsz(j)==Sz1(y)) js0=js0+1; Sz(js0)=xhsz(j); end; end;
end; end
    for(i=1:n) xhsz(i)=Sz(i); end; end
for(x=1:fl) js=0; flsz(x)=0;      %获得分类情况：每一子类的元素个数
    for(i=1:n) pd=1;
        for(y=1:js) if(Sz(y)==i) pd=0; break; end; end
        if(pd)
            if(js==0) y=0; end
            for(j=1:n) if(R(i, j)>=lmd(x)) js=js+1; Sz(js)=j; end; end
            flsz(x)=flsz(x)+1; Sz0(flsz(x))=js-y; end; end
    js0=1;
    for(i=1:flsz(x)) y=1;
        for(j=1:flsz(x))
            if(Sz(y)==xhsz(js0)) flqksz(x, i)=Sz0(j); js0=js0+Sz0(j); break; end
            y=y+Sz0(j); end; end; end
F_result=figure('name', '动态聚类图', 'color', 'w')
axis('off'); kd=30; Gd=40; y=fl*Gd+Gd; lx=80; text(24, y+Gd/2, 'λ');
for(i=1:n)
    text(lx-5+i*Kd-0.4*Kd*(xhsz(i)>9), y+Gd/2, int2str(xhsz(i)));
    line([lx+i*Kd, lx+i*Kd], [y, y-Gd]); linesz(i)=lx+i*Kd; end
text(lx*1.5+i*Kd, y+Gd/2, '分类数'); y=y-Gd;
for(x=1:fl)
    text(8, y-Gd/2, num2str(lmd(x))); js0=1; js1=0;
    if (x==1) for (i=1:flsz(x))
            js1=flqksz(x, i)-1;
            if(js1) line([linesz(js0), linesz(js0+js1)], [y, y]); end
            line([(linesz(js0+js1)+linesz(js0))/2, (linesz(js0+js1)+linesz(js0))/
            2], [y, y-Gd]);
            linesz(i)=(linesz(js0+js1)+linesz(js0))/2;
            js0=js0+js1+1; end
    else for(i=1:flsz(x))
```

```
            js1=js1+flqksz(x, i); js2=0; pd=0;
            for(j=1:flsz(x-1))
                js2=js2+flqksz(x-1, j);
                 if(js2==js1) pd=1; break; end; end
            if(j~=js0) line([linesz(js0) , linesz(j) ], [y, y]); end
            line([(linesz(js0)+linesz(j))/2, (linesz(js0)+linesz(j))/2], [y, y-
           Gd]);
            linesz(i)=(linesz(js0)+linesz(j))/2;
            js0=j+1; end; end
text(2*lx+n*Kd, y-Gd/3, int2str(flsz(x))); y=y-Gd; end
```

附录 E　某市城市基础运行领域的信息组织模糊聚类关键计算数据

表 E.1　基于欧式距离的信息组织模糊相似矩阵 R

00	0.63	0.82	0.47	0.55	0.47	0.85	0.72	0.47	0.47	0.59	0.57	0.47	0.57	0.57	0.57	0.67	0.59	0.66	0.88	0.71	0.53	0.72	0.65
63	1.00	0.77	0.68	0.68	0.68	0.57	0.84	0.68	0.68	0.75	0.79	0.68	0.79	0.79	0.79	0.73	0.79	0.87	0.58	0.85	0.78	0.43	0.76
82	0.77	1.00	0.58	0.64	0.58	0.76	0.85	0.58	0.58	0.68	0.69	0.58	0.69	0.69	0.69	0.75	0.70	0.79	0.79	0.84	0.65	0.59	0.76
47	0.68	0.58	1.00	0.86	1.00	0.48	0.72	0.99	1.00	0.84	0.87	1.00	0.87	0.87	0.87	0.75	0.86	0.74	0.44	0.73	0.89	0.31	0.80
55	0.68	0.64	0.86	1.00	0.86	0.59	0.78	0.86	0.86	0.85	0.85	0.86	0.85	0.85	0.85	0.85	0.86	0.75	0.53	0.76	0.82	0.42	0.84
47	0.68	0.58	1.00	0.86	1.00	0.48	0.72	0.99	1.00	0.84	0.87	1.00	0.87	0.87	0.87	0.75	0.86	0.74	0.44	0.73	0.89	0.31	0.80
85	0.57	0.76	0.48	0.59	0.48	1.00	0.69	0.48	0.48	0.59	0.56	0.48	0.56	0.56	0.56	0.69	0.58	0.61	0.90	0.67	0.52	0.73	0.65
72	0.84	0.85	0.72	0.78	0.72	0.69	1.00	0.72	0.72	0.81	0.83	0.72	0.83	0.83	0.83	0.87	0.84	0.90	0.68	0.94	0.79	0.52	0.87
47	0.68	0.58	0.99	0.86	0.99	0.48	0.72	1.00	0.99	0.84	0.87	0.99	0.87	0.87	0.87	0.75	0.86	0.74	0.43	0.73	0.90	0.31	0.80
47	0.68	0.58	1.00	0.86	1.00	0.48	0.72	0.99	1.00	0.84	0.87	1.00	0.87	0.87	0.87	0.75	0.86	0.74	0.44	0.73	0.89	0.31	0.80
59	0.75	0.68	0.84	0.85	0.84	0.59	0.81	0.84	0.84	1.00	0.86	0.84	0.86	0.86	0.86	0.82	0.88	0.78	0.55	0.80	0.89	0.45	0.82
57	0.79	0.69	0.87	0.85	0.87	0.56	0.83	0.87	0.87	0.86	1.00	0.87	1.00	1.00	1.00	0.82	0.95	0.87	0.53	0.85	0.92	0.39	0.89
47	0.68	0.58	1.00	0.86	1.00	0.48	0.72	0.99	1.00	0.84	0.87	1.00	0.87	0.87	0.87	0.75	0.86	0.74	0.44	0.73	0.89	0.31	0.80
57	0.79	0.69	0.87	0.85	0.87	0.56	0.83	0.87	0.87	0.86	1.00	0.87	1.00	1.00	1.00	0.82	0.95	0.87	0.53	0.85	0.92	0.39	0.89

续表

57	0.79	0.69	0.87	0.85	0.87	0.56	0.83	0.87	0.87	0.86	1.00	0.87	1.00	1.00	1.00	0.82	0.95	0.87	0.53	0.85	0.92	0.39	0.89
57	0.79	0.69	0.87	0.85	0.87	0.56	0.83	0.87	0.87	0.86	1.00	0.87	1.00	1.00	1.00	0.82	0.95	0.87	0.53	0.85	0.92	0.39	0.89
67	0.73	0.75	0.75	0.85	0.75	0.69	0.87	0.75	0.75	0.82	0.82	0.75	0.82	0.82	0.82	1.00	0.85	0.81	0.65	0.83	0.77	0.53	0.85
59	0.79	0.70	0.86	0.86	0.86	0.58	0.84	0.86	0.86	0.88	0.95	0.86	0.95	0.95	0.95	0.85	1.00	0.86	0.54	0.84	0.91	0.42	0.88
66	0.87	0.79	0.74	0.75	0.74	0.61	0.90	0.74	0.74	0.78	0.87	0.74	0.87	0.87	0.87	0.81	0.86	1.00	0.60	0.92	0.81	0.44	0.87
88	0.58	0.79	0.44	0.53	0.44	0.90	0.68	0.43	0.44	0.55	0.53	0.44	0.53	0.53	0.53	0.65	0.54	0.60	1.00	0.66	0.49	0.75	0.61
71	0.85	0.84	0.73	0.76	0.73	0.67	0.94	0.73	0.73	0.80	0.85	0.73	0.85	0.85	0.85	0.83	0.84	0.92	0.66	1.00	0.80	0.49	0.90
53	0.78	0.65	0.89	0.82	0.89	0.52	0.79	0.90	0.89	0.89	0.92	0.89	0.92	0.92	0.92	0.77	0.91	0.81	0.49	0.80	1.00	0.36	0.83
72	0.43	0.59	0.31	0.42	0.31	0.73	0.52	0.31	0.31	0.45	0.39	0.31	0.39	0.39	0.39	0.53	0.42	0.44	0.75	0.49	0.36	1.00	0.45
65	0.76	0.76	0.80	0.84	0.80	0.65	0.87	0.80	0.80	0.82	0.89	0.80	0.89	0.89	0.89	0.85	0.88	0.87	0.61	0.90	0.83	0.45	1.00

表 E.2　基于切比雪夫距离的信息组织模糊相似矩阵 *R*

00	0.62	0.77	0.50	0.50	0.50	0.83	0.75	0.50	0.50	0.50	0.62	0.50	0.62	0.62	0.62	0.67	0.62	0.63	0.88	0.75	0.50	0.62	0.67
62	1.00	0.74	0.50	0.50	0.50	0.50	0.83	0.50	0.50	0.67	0.67	0.50	0.67	0.67	0.67	0.67	0.67	0.83	0.50	0.83	0.67	0.50	0.67
77	0.74	1.00	0.50	0.50	0.50	0.75	0.83	0.50	0.50	0.62	0.67	0.50	0.67	0.67	0.67	0.67	0.67	0.75	0.75	0.83	0.62	0.54	0.67
50	0.50	0.50	1.00	0.87	1.00	0.50	0.67	0.98	1.00	0.83	0.83	1.00	0.83	0.83	0.83	0.75	0.83	0.67	0.50	0.67	0.83	0.50	0.75
50	0.50	0.50	0.87	1.00	0.87	0.62	0.67	0.85	0.87	0.83	0.83	0.87	0.83	0.83	0.83	0.83	0.83	0.67	0.50	0.67	0.83	0.50	0.83
50	0.50	0.50	1.00	0.87	1.00	0.50	0.67	0.98	1.00	0.83	0.83	1.00	0.83	0.83	0.83	0.75	0.83	0.67	0.50	0.67	0.83	0.50	0.75

续表

83	0.50	0.75	0.50	0.62	0.50	1.00	0.63	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.63	0.50	0.50	0.83	0.63	0.50	0.71	0.63
75	0.83	0.83	0.67	0.67	0.67	0.63	1.00	0.67	0.67	0.75	0.83	0.67	0.83	0.83	0.83	0.83	0.83	0.88	0.63	0.90	0.75	0.53	0.83
50	0.50	0.50	0.98	0.85	0.98	0.50	0.67	1.00	0.98	0.83	0.83	0.98	0.83	0.83	0.83	0.75	0.83	0.67	0.50	0.67	0.83	0.50	0.75
50	0.50	0.50	1.00	0.87	1.00	0.50	0.67	0.98	1.00	0.83	0.83	1.00	0.83	0.83	0.83	0.75	0.83	0.67	0.50	0.67	0.83	0.50	0.75
50	0.67	0.62	0.83	0.83	0.83	0.50	0.75	0.83	0.83	1.00	0.88	0.83	0.88	0.88	0.88	0.75	0.88	0.75	0.50	0.75	0.88	0.50	0.75
62	0.67	0.67	0.83	0.83	0.83	0.50	0.83	0.83	0.83	0.88	1.00	0.83	1.00	1.00	1.00	0.78	0.95	0.83	0.50	0.83	0.88	0.50	0.87
50	0.50	0.50	1.00	0.87	1.00	0.50	0.67	0.98	1.00	0.83	0.83	1.00	0.83	0.83	0.83	0.75	0.83	0.67	0.50	0.67	0.83	0.50	0.75
62	0.67	0.67	0.83	0.83	0.83	0.50	0.83	0.83	0.83	0.88	1.00	0.83	1.00	1.00	1.00	0.78	0.95	0.83	0.50	0.83	0.88	0.50	0.87
62	0.67	0.67	0.83	0.83	0.83	0.50	0.83	0.83	0.83	0.88	1.00	0.83	1.00	1.00	1.00	0.78	0.95	0.83	0.50	0.83	0.88	0.50	0.87
62	0.67	0.67	0.83	0.83	0.83	0.50	0.83	0.83	0.83	0.88	1.00	0.83	1.00	1.00	1.00	0.78	0.95	0.83	0.50	0.83	0.88	0.50	0.87
67	0.67	0.67	0.75	0.83	0.75	0.63	0.83	0.75	0.75	0.75	0.78	0.75	0.78	0.78	0.78	1.00	0.83	0.78	0.63	0.78	0.75	0.57	0.76
62	0.67	0.67	0.83	0.83	0.83	0.50	0.83	0.83	0.83	0.88	0.95	0.83	0.95	0.95	0.95	0.83	1.00	0.83	0.50	0.83	0.88	0.50	0.87
63	0.83	0.75	0.67	0.67	0.67	0.50	0.88	0.67	0.67	0.75	0.83	0.67	0.83	0.83	0.83	0.78	0.83	1.00	0.50	0.88	0.75	0.50	0.83
88	0.50	0.75	0.50	0.50	0.50	0.83	0.63	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.63	0.50	0.50	1.00	0.63	0.50	0.68	0.63
75	0.83	0.83	0.67	0.67	0.67	0.63	0.90	0.67	0.67	0.75	0.83	0.67	0.83	0.83	0.83	0.78	0.83	0.88	0.63	1.00	0.75	0.52	0.83
50	0.67	0.62	0.83	0.83	0.83	0.50	0.75	0.83	0.83	0.88	0.88	0.83	0.88	0.88	0.88	0.75	0.88	0.75	0.50	0.75	1.00	0.50	0.75
62	0.50	0.54	0.50	0.50	0.50	0.71	0.53	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.57	0.50	0.50	0.68	0.52	0.50	1.00	0.50
67	0.67	0.67	0.75	0.83	0.75	0.63	0.83	0.75	0.75	0.75	0.87	0.75	0.87	0.87	0.87	0.76	0.87	0.83	0.63	0.83	0.75	0.50	1.00

表 E.3　基于欧式距离的传递闭包 *B*

00	0.82	0.82	0.82	0.82	0.82	0.88	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.88	0.82	0.82	0.75	0.82
82	1.00	0.85	0.87	0.86	0.87	0.82	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.82	0.87	0.87	0.75	0.87
82	0.85	1.00	0.85	0.85	0.85	0.82	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.82	0.85	0.85	0.75	0.85
82	0.87	0.85	1.00	0.86	1.00	0.82	0.89	0.99	1.00	0.89	0.90	1.00	0.90	0.90	0.90	0.87	0.90	0.89	0.82	0.89	0.90	0.75	0.89
82	0.86	0.85	0.86	1.00	0.86	0.82	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.82	0.86	0.86	0.75	0.86
82	0.87	0.85	1.00	0.86	1.00	0.82	0.89	0.99	1.00	0.89	0.90	1.00	0.90	0.90	0.90	0.87	0.90	0.89	0.82	0.89	0.90	0.75	0.89
88	0.82	0.82	0.82	0.82	0.82	1.00	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.90	0.82	0.82	0.75	0.82
82	0.87	0.85	0.89	0.86	0.89	0.82	1.00	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.87	0.89	0.92	0.82	0.94	0.89	0.75	0.90
82	0.87	0.85	0.99	0.86	0.99	0.82	0.89	1.00	0.99	0.89	0.90	0.99	0.90	0.90	0.90	0.87	0.90	0.89	0.82	0.89	0.90	0.75	0.89
82	0.87	0.85	1.00	0.86	1.00	0.82	0.89	0.99	1.00	0.89	0.90	1.00	0.90	0.90	0.90	0.87	0.90	0.89	0.82	0.89	0.90	0.75	0.89
82	0.87	0.85	0.89	0.86	0.89	0.82	0.89	0.89	0.89	1.00	0.89	0.89	0.89	0.89	0.89	0.87	0.89	0.89	0.82	0.89	0.89	0.75	0.89
82	0.87	0.85	0.90	0.86	0.90	0.82	0.89	0.90	0.90	0.89	1.00	0.90	1.00	1.00	1.00	0.87	0.95	0.89	0.82	0.89	0.92	0.75	0.89
82	0.87	0.85	1.00	0.86	1.00	0.82	0.89	0.99	1.00	0.89	0.90	1.00	0.90	0.90	0.90	0.87	0.90	0.89	0.82	0.89	0.90	0.75	0.89
82	0.87	0.85	0.90	0.86	0.90	0.82	0.89	0.90	0.90	0.89	1.00	0.90	1.00	1.00	1.00	0.87	0.95	0.89	0.82	0.89	0.92	0.75	0.89
82	0.87	0.85	0.90	0.86	0.90	0.82	0.89	0.90	0.90	0.89	1.00	0.90	1.00	1.00	1.00	0.87	0.95	0.89	0.82	0.89	0.92	0.75	0.89
82	0.87	0.85	0.90	0.86	0.90	0.82	0.89	0.90	0.90	0.89	1.00	0.90	1.00	1.00	1.00	0.87	0.95	0.89	0.82	0.89	0.92	0.75	0.89
82	0.87	0.85	0.87	0.86	0.87	0.82	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	1.00	0.87	0.87	0.82	0.87	0.87	0.75	0.87

续表

82	0.87	0.85	0.90	0.86	0.90	0.82	0.89	0.90	0.90	0.89	0.95	0.90	0.95	0.95	0.95	0.87	1.00	0.89	0.82	0.89	0.92	0.75	0.89
82	0.87	0.85	0.89	0.86	0.89	0.82	0.92	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.87	0.89	1.00	0.82	0.92	0.89	0.75	0.90
88	0.82	0.82	0.82	0.82	0.82	0.90	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	1.00	0.82	0.82	0.75	0.82
82	0.87	0.85	0.89	0.86	0.89	0.82	0.94	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.87	0.89	0.92	0.82	1.00	0.89	0.75	0.90
82	0.87	0.85	0.90	0.86	0.90	0.82	0.89	0.90	0.90	0.89	0.92	0.90	0.92	0.92	0.92	0.87	0.92	0.89	0.82	0.89	1.00	0.75	0.89
75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	1.00	0.75
82	0.87	0.85	0.89	0.86	0.89	0.82	0.90	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.87	0.89	0.90	0.82	0.90	0.89	0.75	1.00

表 E.4　基于切比雪夫距离的传递闭包 *B*

00	0.77	0.77	0.77	0.77	0.77	0.83	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.88	0.77	0.77	0.71	0.77
77	1.00	0.83	0.83	0.83	0.83	0.77	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.77	0.83	0.83	0.71	0.83
77	0.83	1.00	0.83	0.83	0.83	0.77	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.77	0.83	0.83	0.71	0.83
77	0.83	0.83	1.00	0.87	1.00	0.77	0.83	0.98	1.00	0.83	0.83	1.00	0.83	0.83	0.83	0.83	0.83	0.83	0.77	0.83	0.83	0.71	0.83
77	0.83	0.83	0.87	1.00	0.87	0.77	0.83	0.87	0.87	0.83	0.83	0.87	0.83	0.83	0.83	0.83	0.83	0.83	0.77	0.83	0.83	0.71	0.83
77	0.83	0.83	1.00	0.87	1.00	0.77	0.83	0.98	1.00	0.83	0.83	1.00	0.83	0.83	0.83	0.83	0.83	0.83	0.77	0.83	0.83	0.71	0.83
83	0.77	0.77	0.77	0.77	0.77	1.00	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.83	0.77	0.77	0.71	0.77
77	0.83	0.83	0.83	0.83	0.83	0.77	1.00	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.88	0.77	0.90	0.83	0.71	0.83
77	0.83	0.83	0.98	0.87	0.98	0.77	0.83	1.00	0.98	0.83	0.83	0.98	0.83	0.83	0.83	0.83	0.83	0.83	0.77	0.83	0.83	0.71	0.83

续表

77	0.83	0.83	1.00	0.87	1.00	0.77	0.83	0.98	1.00	0.83	0.83	1.00	0.83	0.83	0.83	0.83	0.83	0.83	0.77	0.83	0.83	0.71	0.83
77	0.83	0.83	0.83	0.83	0.83	0.77	0.83	0.83	0.83	1.00	0.88	0.83	0.88	0.88	0.88	0.83	0.88	0.83	0.77	0.83	0.88	0.71	0.87
77	0.83	0.83	0.83	0.83	0.83	0.77	0.83	0.83	0.83	0.88	1.00	0.83	1.00	1.00	1.00	0.83	0.95	0.83	0.77	0.83	0.88	0.71	0.87
77	0.83	0.83	1.00	0.87	1.00	0.77	0.83	0.98	1.00	0.83	0.83	1.00	0.83	0.83	0.83	0.83	0.83	0.83	0.77	0.83	0.83	0.71	0.83
77	0.83	0.83	0.83	0.83	0.83	0.77	0.83	0.83	0.83	0.88	1.00	0.83	1.00	1.00	1.00	0.83	0.95	0.83	0.77	0.83	0.88	0.71	0.87
77	0.83	0.83	0.83	0.83	0.83	0.77	0.83	0.83	0.83	0.88	1.00	0.83	1.00	1.00	1.00	0.83	0.95	0.83	0.77	0.83	0.88	0.71	0.87
77	0.83	0.83	0.83	0.83	0.83	0.77	0.83	0.83	0.83	0.88	1.00	0.83	1.00	1.00	1.00	0.83	0.95	0.83	0.77	0.83	0.88	0.71	0.87
77	0.83	0.83	0.83	0.83	0.83	0.77	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	1.00	0.83	0.83	0.77	0.83	0.83	0.71	0.83
77	0.83	0.83	0.83	0.83	0.83	0.77	0.83	0.83	0.83	0.88	0.95	0.83	0.95	0.95	0.95	0.83	1.00	0.83	0.77	0.83	0.88	0.71	0.87
77	0.83	0.83	0.83	0.83	0.83	0.77	0.88	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	1.00	0.77	0.88	0.83	0.71	0.83
88	0.77	0.77	0.77	0.77	0.77	0.83	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	1.00	0.77	0.77	0.71	0.77
77	0.83	0.83	0.83	0.83	0.83	0.77	0.90	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.88	0.77	1.00	0.83	0.71	0.83
77	0.83	0.83	0.83	0.83	0.83	0.77	0.83	0.83	0.83	0.88	0.88	0.83	0.88	0.88	0.88	0.83	0.88	0.83	0.77	0.83	1.00	0.71	0.87
71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	1.00	0.71
77	0.83	0.83	0.83	0.83	0.83	0.77	0.83	0.83	0.83	0.87	0.87	0.83	0.87	0.87	0.87	0.83	0.87	0.83	0.77	0.83	0.87	0.71	1.00

主要外文名词缩写索引

A

ADSL：Asymmetrical Digital Subscriber Loop，非对称数字用户环路

ANSI：American National Standards Institute，美国国家标准学会

B

BSI：Britain Standards Institute，英国标准协会

C

CCSA：China Communications Standards Association，中国通信标准化协会

CDMA：Code Division Multiple Access，码分多址

CEN：Comité Européen de Normalization［法］，欧洲标准委员会

CENLEC：European Committee for Electrotechnical Standardization，欧洲电工标准化委员会

CPE：Customer Premise Equipment，用户前置设备

CRC：Cyclic Redundancy Check，循环冗余校验

D

DAS：Direct Attached Storage，直连存储

DIN：Deutsches Institut für Normung［德］，德国标准化协会

DKE：德国电气电工信息技术委员会

DM：Data Mining，数据挖掘

DSS：Decision Support System，决策支持系统

DSSO：　Decision Support System Optimizer，决策支持系统优化器

E

EFA FTI：Education for All-Fast Track Initiative，全民教育-快速跟踪计划

EPC：　Electronic Product Code，产品电子代码

ETSI：　European Telecommunications Standards Institute，欧洲电信标准化协会

G

GDSS：　Group Decision Support System，群体决策支持系统

GPE：　Global Partnership for Education，全球教育伙伴组织

GPRS：　General Packet Radio Service，通用无线分组业务

GPS：　Global Positioning System，全球定位系统

GS1：　Globe Standard 1，国际物品编码协会



H

HA：　High Availability，高可用性

HDFS：　Hadoop Distributed File System，Hadoop 分布式文件系统

HDI：　Health Data Initiative，健康卫生数据行动计划

HTTP：　Hypertext Transfer Protocol，超文本传输协议

I

IDC：　Internet Data Center，互联网数据中心

IEC：　International Electrotechnical Commission，国际电工委员会

IEEE：　Institute of Electrical and Electronics Engineers，美国电气和电子工程师协会

IOT：　Internet of Things，物联网

ISO：International Organization for Standardization，国际标准化组织

ITU：International Telecommunication Union，国际电信联盟

ITU-T：International Telecommunication Union-Telecommunication Sector，国际电信联盟标准化部

J

JMS：Java Message Service，Java 消息服务

K

KDD：Knowledge Discovery in Database，知识发现

M

MCU：Microcontroller Unit，微控制器

MIMO：Multiple-Input Multiple-Output，多输入多输出

MOOC：Massive Open Online Courses，大型开放式在线课程

M2M：Machine to Machine，机-机

N

NAS：Network Attached Storage，网络附加存储

O

OECD：Organization for Economic Cooperation and Development，经济合作与发展组织

OFDM：Orthogonal Frequency Division Multiplexing，正交频分复用

OLAP：Online Analytical Processing，联机分析处理

OS：Operating System，操作系统

P

PSTN：Public Switched Telephone Network，公共电话交换网

R

RDF：Resource Description Framework，资源描述框架

RFID：Radio Frequency Identification，射频识别

S

SAC：Standardization Administration of China，国家标准化管理委员会

SNA：Social Network Analysis，社会网络分析

SOA：Service-Oriented Architecture，面向服务的体系结构

U

URI：Uniform Resource Identifier，统一资源标识符

3G：The 3rd-Generation Mobile Communication Technology，第三代移动通信技术

3GPP：The 3rd Generation Partnership Project，第三代合作伙伴计划

4G：The 4th-Generation Mobile Communication Technology，第四代移动通信技术

参 考 文 献

[1] Agrawal R,JE Gehrke,D Gunopulos,et al. Automatic subspace clustering of high dimensional data for data mining applications. Proceedings of the ACM-SIGMOD'98 Int. Conf. Management of Data. ACM Press,1998,94-105.

[2] Ahmed A. ,E. P. Xing. Recovering time-varying networks of dependencies in social and biological studies. Proc Natl Acad Sci USA,2009,106(29):11878-11883.

[3] Alberto D R A ,Steven A D . An Approach to Facilitate Security Assurance for Information Sharing and Exchange in Big-Data Applications [J]. Emerging Trends in ICT Security,2014:65-83.

[4] Al-Harbi SH, VJ Rayward-Smith. Adapting k-means for supervised clustering. Applied Intelligence,2006,24(3):219-226.

[5] Amadou Boubacar Habiboulaye,Stéphane Lecoeuche,Salah Maouche. SAKM:Self-adaptive kernel machine A kernel-based algorithm for online clustering. Neural Networks,2008,21(9):1287-1301.

[6] Ankerst Mihael,Markus M. Breunig,Hans-Peter Kriegel,et al. OPTICS:ordering points to identify the clustering structure. Proceedings of the ACM SIGMOD International Conference on Management of Data. ACM Press,1999,49-60.

[7] Apostolico A. ,O. Denas. Fast Algorithms for Computing Sequence Distances by Exhaustive Substring Composition. Algorithms Mol Biol,2008,3(1):13.

[8] Arifin Agus Zainal, Akira Asano. Image segmentation by histogram thresholding using hierarchical cluster analysis. Pattern Recognition Letters,2006,27(13):1515-1521.

[9] Bagirov Adil M. Modified global k-means algorithm for minimum sum-of-squares clustering problems. Pattern Recognition,2008,41(10):3192-3199.

[10] Balan,R. K. Nguyen,K. X. ,Jiang,L. Real-time trip information service for a large taxi fleet[C]. Proceedings of the 9th International Conference on Mobile Systems,Applications,and Services,2011,99-112.



[11] Bandyopadhyay Sanghamitra,Sriparna Saha. GAPS: A clustering method using a new point symmetry-based distance measure. Pattern Recognition,2007,40(12):3430-3451.

[12] Baraldi A., P. Blonda. A survey of fuzzy clustering algorithms for pattern recognition. I. IEEE Trans Syst Man Cybern B Cybern,1999,29(6):778-785.

[13] Baraldi A., P. Blonda. A survey of fuzzy clustering algorithms for pattern recognition. II. IEEE Trans Syst Man Cybern B Cybern,1999,29(6):786-801.

[14] Beñat Bilbao-Osorio, Soumitra Dutta, Bruno Lanvin. The Global Information Technology Report 2014[R]. 2014 World Economic Forum,2014.

[15] Bezdek JC. Cluster validity with fuzzy sets. Journal of Cybernetics,1974,3(3):58-73.

[16] Bicego Manuele,Mario A. T. Figueiredo. Soft clustering using weighted one-class support vector machines. Pattern Recognition,2009,42(1):27-32.

[17] Boyd Cohen. The Top 10 Smart Cities On The Planet [EB/OL]. http://www.fastcoexist. com/1679127/the-top-10-smart-cities-on-the-planet,2012/2013-03-08.

[18] Breiger R L,Boorman S A,Arabie P. An algorithm for clustering relational data with applications to social network analysis and comparison with multidimensional scaling[J]. Journal of Mathematical Psychology,1975,12(3):328-383.

[19] Camastra Francesco, Alessandro Verri. A novel Kernel method for clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence,2005,27(5):801-805.

[20] Caragliu A,Del Bo C,Nijkamp P. Smart cities in Europe(2009) [R/OL]. [2011-03-01]. http://www.cers.tuke.sk/cers2009/PDF/01_03_Nijkamp.pdf.

[21] Caragliu,A. Del Bo,C. and Nijkamp,P. Smart cities in Europe [J]. Faculty of Economics,Business Administration and Econometrics,2009:131-143.

[22] Chen C H,Chen J A. Interactive diagnostic plots for multidimensional scaling with applications in psychosis disorder data analysis[J]. Statistica Sinica,2000,10:665-691.

[23] Chen C H. Generalized association plots: information visualization iteratively

generated correlation matrices[J]. Statitica Sinica,2002,12: 7-29.

[24] Cheung C, Black J. A Reappraisal of the Intervening Opportunities Model of Commuter Behaviour[J]. Road and Transport Research,2008,17(2) : 3-18.

[25] Chung Kuo-Liang,Jhin-Sian Lin. Faster and more robust point symmetry-based K-means algorithm. Pattern Recognition,2007,40(2): 410-422.

[26] Colin Tankard,Digital Pathways. Big data security [J]. Network Security,2012, 07: 5-8.

[27] Das Swagatam, Ajith Abraham, Amit Konar. Automatic kernel clustering with a Multi-Elitist Particle Swarm Optimization Algorithm. Pattern Recognition Letters,2008,29(5): 688-699.

[28] de Carvalho Francisco de A. T. ,Renata M. C. R. de Souza,Marie Chavent,et al. Adaptive Hausdorff distances and dynamic clustering of symbolic interval data. Pattern Recognition Letters,2006,27(3): 167-179.

[29] Dembele D. ,P. Kastner. Fuzzy C-means method for clustering microarray data. Bioinformatics,2003,19(8): 973-980.

[30] Dhillon Inderjit S. , Yuqiang Guan, Brian Kulis. Weighted graph cuts without eigenvectors a multilevel approach. IEEE Transactions on Pattern Analysis and Machine Intelligence,2007,29(11): 1944-1957.

[31] Du Z. ,Y. Wang,Z. Ji. PK-means: A new algorithm for gene clustering. Comput Biol Chem,2008,32(4): 243-247.

[32] Ester M,H P Kriegel,J Sander. A density-based algorithm for discovering clusters in large spatial databases. Knowledge Discovery and Data Mining (KDD'96). AAAI Press,1996,226-231.

[33] Everett M G,Borgatti S. A testing example for positional analysis techniques[J]. Social Networks,2009(12): 253-260.

[34] Fei Wang,Dexian Zhang, Na Bao. Fuzzy document clustering based on ant colony algorithm [C]. Proceedings of the 6th International Symposium on Neural Networks: Advances in Neural Networks-Part Ⅱ,2009: 709-716.

[35] Filippone Maurizio, Francesco Camastra, Francesco Masulli, et al. A survey of

kernel and spectral methods for clustering. Pattern Recognition, 2008, 41(1): 176-19.

[36] Fortunato S., M. Barthelemy. Resolution limit in community detection. Proc Natl Acad Sci USA, 2007, 104(1): 36-41.

[37] Gan G., J. Wu. A convergence theorem for the fuzzy subspace clustering (FSC) algorithm. Pattern Recognition, 2008, 41(6): 1939-1947.

[38] Gath I, AB Geva. Unsupervised optimal fuzzy clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1989, 11(7): 773-780.

[39] Ghada E, Mustafa Y E, Saleh E, Mohamed S A: Service Oriented Data Integration based on Map Reduce[J]. Alexandria Engineering Journal, 2013, 52(3): 313-318.

[40] Giffinger R, Fertner C, Kramar H, et al. Smart cities Ranking of European medium-sized cities[R]. Centre of Regional Sci-ence, Vienna UT, October 2007.

[41] Girvan M., M. E. Newman. Community structure in social and biological networks. Proc Natl Acad Sci USA, 2002, 99(12): 7821-7826.

[42] Goldberger Jacob, Tamir Tassa. A hierarchical clustering algorithm based on the Hungarian method. Pattern Recognition Letters, 2008, 29(11): 1632-1638.

[43] Grira Nizar, Michel Crucianu, Nozha Boujemaa. Active semi-supervised fuzzy clustering. Pattern Recognition, 2008, 41(5): 1834-1844.

[44] Guha Sudipto, Rajeev Rastogi, Kyuseok Shim. CURE: An efficient clustering algorithm for large databases. SIGMOD Record (ACM Special Interest Group on Management of Data). Croatian Soc Chem Eng, Zagreb, Croatia, 1998, 73-84.

[45] Guha Sudipto, Rajeev Rastogi, Kyuseok Shim. Rock: a robust clustering algorithm for categorical attributes. Information Systems, 2000, 25(5): 345-366.

[46] Hamerly G., C. Elkan. Learning the k in k-means. Advances in Neural Information Processing Systems 16. MIT Press, 2003, 281-289.

[47] Han J W, Micheline K. Data Mining: Concepts and Techniques[M]. Beijing: High Education Press, 2001.

[48] Handl Julia, Joshua Knowles, Douglas B. Kell. Computational cluster validation in post-genomic data analysis. Bioinformatics, 2005, 21(15): 3201-3212.

[49] Harrison, C. Eckman, B. Hamilton, et al. Foundations for Smarter Cities [J]. IBM Journal of Research and Development, 54(4).

[50] Hathaway Richard J., James C. Bezdek. Extending fuzzy and probabilistic clustering to very large data sets. Computational Statistics and Data Analysis, 2006, 51(1): 215-234.

[51] Hinneburg A, DA Keim. An efficient approach to clustering in large multimedia databases with noise. Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining KDD'98. AAAI Press, 1998, 58-65.

[52] Huang Joshua Zhexue, Michael K. Ng, Hongqiang Rong, et al. Automated variable weighting in k-means type clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(5): 657-668.

[53] Huntsberger TL, P Ajjimarangsee. Parallel self-organizing feature maps for unsupervised pattern recognition. International Journal of general systems, 1990, 16(4): 357-372.

[54] Izakian H, Abraham A. Fuzzy C-means and fuzzy swarm for fuzzy clustering problem[J]. Expert Systems with Applications, 2011, 38(3): 1835-1838.

[55] Jain A. K., M. N. Murty. Data clustering : a review. ACM Computing Surveys, 1999, 31(3): 264-323.

[56] James B, Dan B. Cooperation through imitation [J]. Games and Economic Behavior, 2009, 8(2): 15-19.

[57] Jamias S B. The adoption of rice-mungbean technology in two pangasinam villages: a black modeling analysis[J]. Journal of Cropping Science, 2007, 22(2): 89-98.

[58] Jeong H., B. Tombor, R. Albert, et al. The large-scale organization of metabolic networks. Nature, 2000, 407(6804): 651-654.

[59] Kaplan N., M. Friedlich, M. Fromer, et al. A functional hierarchical organization of the protein sequence space. BMC Bioinformatics, 2004, 5: 196.

[60] Karypis George, Eui-Hong Han, Vipin Kumar. Chameleon: Hierarchical clustering using dynamic modeling. Computer, 1999, 32(8): 68-75.



[61] Kim Dae-Won, KiYoung Lee, Doheon Lee, et al. A kernel-based subtractive clustering method. Pattern Recognition Letters,2005,26(7): 879-891.

[62] Kocsor A.,A. Kertesz-Farkas,L. Kajan,et al. Application of compression-based distance measures to protein sequence classification: a methodological study. Bioinformatics,2006,22(4): 407-412.

[63] Krause A.,J. Stoye,M. Vingron. Large scale hierarchical clustering of protein sequences. BMC Bioinformatics,2005,6: 15.

[64] Krish K. Integration of Big Data and Data Warehousing [J]. Data Warehousing in the Age of Big Data,2013: 199-217.

[65] Lai Jim Z. C.,Yi-Ching Liaw,Julie Liu. Fast k-nearest-neighbor search based on projection and triangular inequality. Pattern Recognition,2007,40(2): 351-359.

[66] Laskaris Nikolaos A.,Stefanos P. Zafeiriou. Beyond FCM: Graph-theoretic post-processing algorithms for learning and representing the data structure. Pattern Recognition,2008,41(8): 2630-2644.

[67] Lee Sang Wan,Yong Soo Kim,Kwang-Hyun Park,et al. Iterative Bayesian fuzzy clustering toward flexible icon-based assistive software for the disabled. Information Sciences,2010,180(3): 325-340.

[68] Li Chaoshun,Zhou Jianzhong,Kou Pangao,et al. A novel chaotic particle swarm optimization based fuzzy clustering algorithm[J]. Neurocomputing,2012,83(4): 98-109.

[69] Li Y, Yu F. A new validity function for fuzzy clustering[J]. International Conference on Computational Intelligence and Natural Computing, 2009, 1: 462-465.

[70] Liao Liang,Tusheng Lin,Bi Li. MRI brain image segmentation and bias field correction based on fast spatially constrained kernel clustering approach. Pattern Recognition Letters,2008,29(10): 1580-1588.

[71] Ling Haibin,David W. Jacobs. Shape classification using the inner-distance. IEEE Transactions on Pattern Analysis and Machine Intelligence,2007,29(2): 286-299.

[72] Liou J J H,Yen L,Tzeng G H. Building an effective safety management system for

airlines[J]. Journal of Air Transport Management,2008,14(1):20-26.

[73] Liu Rujie, Yuehong Wang, Takayuki Baba, et al. SVM-based active feedback in image retrieval using clustering and unlabeled data. Pattern Recognition,2008,41(8):2645-2655.

[74] Liu, S. Towards mobility-based clustering [C]. Proceedings of the 16th ACMSIGKDD International Conference on Knowledge Discovery and Data Mining, 2010,919-928.

[75] Loewenstein Yaniv, Elon Portugaly, Menachem Fromer, et al. Efficient algorithms for accurate hierarchical clustering of huge datasets: tackling the entire protein space. Bioinformatics,2008,24(13):i41-i49.

[76] Luxburg U Von, O Bousquet, M Belkin. Limits of spectral clustering. Advances in neural information processing systems 17. MIT Press,2005,1020-1027.

[77] Luxburg Ulrike. A tutorial on spectral clustering. Statistics and Computing,2007, 17(4):395-416.

[78] Ma Jianmin, Minh N. Nguyen, Jagath C. Rajapakse. Gene Classification Using Codon Usage and Support Vector Machines. Computational Biology and Bioinformatics, IEEE/ACM Transactions on,2009,6(1):134-143.

[79] Ma T C. Network analysis in policymaking: the case of significant labor events in Taiwan[J]. Issues and Studies,2008,44(1):133-168.

[80] Martinetz T. M. ,S. G. Berkovich, K. J. Schulten. Neural-gas network for vector quantization and its application to time-series prediction. IEEE Trans Neural Netw,1993,4(4):558-569.

[81] Maslov S. ,K. Sneppen. Specificity and stability in topology of protein networks. Science,2002,296(5569):910-913.

[82] Masson Marie-Hélène, T. Denoeux. ECM: An evidential version of the fuzzy c-means algorithm. Pattern Recognition,2008,41(4):1384-1397.

[83] McCane Brendan, Michael Albert. Distance functions for categorical and mixed variables. Pattern Recognition Letters,2008,29(7):986-993.

[84] McQuitty L L. Multiple clusters, types, and dimensions from iterative



intercolumnar correlational analysis [J]. Multivariate Behavioral Research, 1968, 3 (4): 465-477.

[85] Ng A, M Jordan, Y Weiss. On Spectral Clustering: Analysis and an algorithm. Advances in neural information processing systems. MIT Press, 2002, 849-856.

[86] Noebert A. Streitz. Smart Cities, Ambient Intelligence and Universal Access[J]. 2011, HCI(7), 425-432.

[87] Ostling Annette, John Harte, Jessica Green, et al. Self-Similarity and Clustering in the Spatial Distribution of Species. Science, 2000, 290(5492): 671a.

[88] P. C. Wong, H.-W. Shen, C. R. Johnson, et al. The Top 10 Challenges in Extreme-Scale Visual Analytics [J]. IEEE CG&A, 2012: 197-207.

[89] Pal NR, K Pal, JM Keller, et al. A possibilistic fuzzy c-means clustering algorithm. IEEE Transactions on Fuzzy Systems, 2005, 13(4): 517-530.

[90] Prettejohn B J, Berryman M J, Mcdonnell M D. Methods for Generating Complex Networks with Selected Structural Properties for Simulations: A Review and Tutorial for Neuroscientists [J]. Frontiersincomputational neuroscience2011, 5 (11) : 1-18.

[91] Qin A. K. , P. N. Suganthan. Enhanced neural gas network for prototype-based clustering. Pattern Recognition, 2005, 38(8): 1275-1288.

[92] Rudolf Giffinger etc. Smart cities-ranking of European medium-sized cities [J]. Centre of Regional Sci-ence, Vienna UT, October 2007.

[93] Sander J, M Ester, HP Kriegel, et al. Density-based clustering in spatial databases: The algorithm gdbscan and its applications. Data mining and knowledge discovery, 1998, 2(2): 169-194.

[94] Schaffers H, Komninos N, Pallot M, et al, Oliveira A. Smart Cities and the Future Internet: Towards Cooperation Frame-works for Innovation[C]//Domingue J, et al. The Future Internet. Heidelberg: Springer 2011: 431-446.

[95] Simone C, Andrea D M. Accessibility and Complex Network Analysis of the U. S. commuting system [J]. Cities, 2013(30): 6-21.

[96] Scott, J. , Social Network Analysis: A Handbook [M]. Sage Publications, 2000.

[97] Sonnhammer E. L. ,V. Hollich. Scoredist: a simple and robust protein sequence distance estimator. Bioinformatics,2005,6: 108.

[98] Staiano A, R Tagliaferri, W Pedrycz. Improving RBF networks performance in regression tasks by means of a supervised fuzzy clustering. Neurocomputing, 2006,69(13-15): 1570-1581.

[99] Sun Da-wei,Chang Gui-ran,Jin Li-zhong,et al. Optimizing grid resource allocation by combining fuzzy clustering with application preference[C]. Proceedings of the 2nd IEEE International Conference on Advanced Computer Control. 2010: 22-27.

[100] Thomas H. Davenport. Competing on Analytics: the New Science of Winning [M]. Boston: Harvard Business School Press,2007.

[101] Tushir Meena, Smriti Srivastava. A new Kernelized hybrid c-mean clustering model with optimized parameters. Applied Soft Computing, 2010, 10 (2): 381-389.

[102] Urban Computing [EB/OL]. http: //research. microsoft. com/en-us/projects/urban-computing,2012.

[103] US Dept. of Energy. ASCR Research: Scientific Discovery through Advanced Computing (SciDAC),2012.

[104] Van Noorden R. Data-sharing: Everything on display [J]. Nature, 2013, 500 (7461): 243-245.

[105] Ververidis D, C Kotropoulos. Information Loss of the Mahalanobis Distance in High Dimensions: Application to Feature Selection. IEEE Transactions on Pattern Analysis and Machine Intelligence,2009,31(12): 2275-2281.

[106] Vijaya P. A. , M. Narasimha Murty, D. K. Subramanian. Efficient bottom-up hybrid hierarchical clustering techniques for protein sequence classification. Pattern Recognition,2006,39(12): 2344-2355.

[107] Vinga S. , R. Gouveia-Oliveira, J. S. Almeida. Comparative evaluation of word composition distances for the recognition of SCOP relationships. Bioinformatics, 2004,20(2): 206-215.

[108] Wang H. , H. Zheng, F. Azuaje. Poisson-based self-organizing feature maps and

hierarchical clustering for serial analysis of gene expression data. IEEE/ACM Trans Comput Biol Bioinform, 2007, 4(2): 163-175.

[109] Wang W, J Yang, R Muntz. STING: A statistical information grid approach to spatial data mining. Proc. 23rd Int. Conf. on Very Large Data Bases. IEEE Press, 1997, 186-195.

[110] Washburn D, Sindhu U. Helping CIOs Understand Smart City Initiatives[R]. Forrester Research, 2010.

[111] Wellman, B., Berkowitz, S. D., Social Structures: A Network Approach[M]. Cambridge, England: Cambridge University Press, 1988.

[112] Winters-Hilt S., S. Merat. SVM clustering. BMC Bioinformatics, 2007, 8(Suppl 7): S18.

[113] Wu F. X. Genetic weighted k-means algorithm for clustering large-scale gene expression data. BMC Bioinformatics, 2008, 9(Suppl 6): S12.

[114] Wu Kuo-Ping, Sheng-De Wang. Choosing the kernel parameters for support vector machines by the inter-cluster distance in the feature space. Pattern Recognition, 2009, 42(5): 710-717.

[115] Xiong H., J. Wu, J. Chen. K-means clustering versus validation measures: a data-distribution perspective. IEEE Trans Syst Man Cybern B Cybern, 2009, 39(2): 318-331.

[116] Yang Xinshe. Firefly algorithms for multimodal optimization[C]. Proc of the 5th International Conference on Stochastic Algorithms: Foundations and Applications. Berlin: Springer-Verlag, 2009: 169-178.

[117] Yu Jie, Jaume Amores, Nicu Sebe, et al. Distance learning for similarity estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(3): 451-462.

[118] Yu Stella X., Jianbo Shi. Multiclass spectral clustering. Proceedings of the IEEE International Conference on Computer Vision. IEEE Press, 2003, 313-319.

[119] Zelnik-Manor L., P. Perona. Self-tuning spectral clustering. Advances in Neural Information Processing Systems 16. MIT Press, 2004, 1601-1608.

[120] Zhang Tian, Raghu Ramakrishnan, Miron Livny. BIRCH: An efficient data clustering method for very large databases. SIGMOD Record (ACM Special Interest Group on Management of Data). ACM,1996,103-114.

[121] Zheng, Y. Urban computing with taxicabs [C]. Proceedings of the 13th International Conference on Ubiquitous Computing,2011,89-98.

[122] Zhu L,Chung F L,Wang S. Generalized fuzzy C-means clustering algorithm with improved fuzzy partitions[J]. IEEE. Trans.,Syst.,Man,Cybern.,2009,39(3):578-591.

[123] 鲍常勇. 我国286个地级及以上城市流动人口分布特征分析[J]. 人口研究. 2007,31(6):67-75.

[124] 曹建君等. 基于信息融合理论的省情信息融合研究[J]. 遥感技术与应用. 2006,21(4):368-371.

[125] 陈金海. 关于情报信息融合处理方法的研究[J]. 情报杂志. 2003(3):63-64.

[126] 陈康,郑纬民. 云计算-系统实例与研究现状[J]. 软件学报. 2009(5).

[127] 陈立,李春香,李志勇. 基于物联网的智慧城市的内涵、特征与要素构成[J]. 硅谷. 2012,(9):15-16.

[128] 陈柳钦. 智慧城市:全球城市发展新热点[J]. 青岛科技大学学报(社会科学版). 2011(3).

[129] 陈铭,王乾晨,张晓海,张晓伟. "智慧城市"评价指标体系研究—以"智慧南京"建设为例[J]. 城市发展研究. 2011,18(5):84-89.

[130] 陈鹏慧,吴宝明. 信息融合技术及其在医疗监护系统中的应用[J]. 国外医学. 生物医学工程分册. 2002(6):249-252.

[131] 陈锐,周永根,沈华,徐浩然. 技术变革与技术标准协同发展的战略思考[J]. 科学学研究. 2013,31(7):1006-1012.

[132] 陈锐. 物联网——后IP时代国家创新发展的重大战略机遇[J]. 中国科学院院刊. 2010,25(1):41-49.

[133] 仇保兴. 智慧地进行城镇建设 积极促进中国城镇可持续发展[J]. 城市发展研究. 2012,21(10):8-10.

[134] 邓贤峰,张晓伟. 城市"智慧化"发展的趋势研究[J]. 电子政务. 2011(4).



[135] 邓贤峰."智慧城市"评价指标体系研究[J]. 发展研究.2010,(12):111-116.

[136] 董世龙,陈宁江,谭瑛等. 面向云环境的集群资源模糊聚类划分算法的优化[J]. 计算机科学.2014,41(9):104-109.

[137] 段平忠. 中国省际人口迁移对经济增长动态收敛的影响[J]. 中国人口·资源与环境.2011,18(12):146-152.

[138] 高丰. 开放数据:概念、现状与机遇[J]. 大数据.2015,014.

[139] 工业和信息化部电信研究院,政府在云计算发展中的作用[J]. 数据通信.2012(8).

[140] 辜胜阻,王敏. 智慧城市建设的理论思考与战略选择[J]. 中国人口·资源与环境.2012,22(5):74-80.

[141] 顾德道,乔雯. 我国智慧城市评价指标体系的构建研究[J]. 未来与发展.2012,35(10):79-83.

[142] 侯贺平. 基于改进辐射模型的乡镇人口流动网络研究[J]. 中国人口.资源与环境.2013,23(8):107-114.

[143] 侯赟慧,刘志彪,岳中刚. 长三角区域经济一体化进程的社会网络分析[J]. 中国软科学.2009(12):90-101.

[144] 胡蓓,王聪颖.基于信息融合的发展中国家高技术产业集群知识融合与创新模型研究[J].图书情报工作.2009,53(2):38-41,73.

[145] 胡小明. 从数字城市到智慧城市资源观念的演变[J]. 电子政务.2011,(8):47-56.

[146] 化柏林. 多源信息融合方法研究[J]. 情报理论与实践.2013,36(11):16-19.

[147] 黄成泉,王士同,蒋亦樟. 熵指数约束的模糊聚类新算法[J]. 计算机研究与发展.2014,51(9):2117-2129.

[148] 纪韶,朱志胜. 中国城市群人口流动与区域经济发展平衡性研究——基于全国第六次人口普查长表数据的分析[J]. 经济理论与经济管理.2014,23(2):5-16.

[149] 雷小锋,谢昆青,林帆,等. 一种基于 K-Means 局部最优性的高效聚类算法. 软件学报.2008,19(7):1683-1692.

[150] 李伯华. 中国流动人口生存发展状况报告——基于重点地区流动人口监测试点调查[J]. 人口研究.2010,34(1):6-18.



[151] 李广建，杨林. 大数据视角下的情报研究与情报研究技术[J]. 图书与情报. 2012(6)：1-8.

[152] 李国杰，程学旗. 大数据研究：未来科技及经济社会发展的重大战略领域[J]. 中国科学院院刊. 2012，27(6)：647-657.

[153] 李国杰. 大数据研究的科学价值[J]. 中国计算机学会通讯. 2012，08(9)：8-15.

[154] 李健，张春梅，李海花. 智慧城市及其评价指标和评估方法研究[J]. 电信网技术. 2012，(1)：1-5.

[155] 李文娟，张启飞，平玲娣等. 基于模糊聚类的云任务调度算法[J]. 通信学报. 2012，33(3)：146-154.

[156] 李贤毅，邓晓宇. 智慧城市评价指标体系研究[J]. 电信网技术. 2011，(10)：43-47.

[157] 励娜，尹怀庭. 我国城乡人口流动的驱动因素分析[J]. 西北大学学报(自然科学版). 2008，38(6)：1019-1023.

[158] 林佳莹，蔡敏智. 台北地区小学教育资源分布地位之探讨：社会网络结构地位分析之应用[J]. 社会与教育研究(台). 2006(6)：71-106.



[159] 林睦纲，刘芳菊，童小娇. 一种基于萤火虫算法的模糊聚类方法[J]. 计算机工程与应用. 2014，50(21)：35-38，73.

[160] 刘军. 法村社会支持网络的整体结构研究块模型及其应用[J]. 社会. 2006，26(3)：69-80.

[161] 刘军. 社会网络分析导论[M]. 北京：社会科学文献出版社. 2004.

[162] 刘军. 整体网分析讲义——UCINET 软件实用指南[M]. 上海：上海人民出版社. 2009.

[163] 刘明香. 信息融合技术在知识信息化中的应用[J]. 情报杂志. 2001(40)：23.

[164] 刘平峰，章佩璐，张军. 面向主题的 Web 信息融合模型[J]. 图书情报工作. 2011，55(8)：40-43.

[165] 刘钟美，贡金涛. 基于社会网络分析的美国 ILS 图书情报院校关系研究[J]. 农业图书情报学刊(台). 2010，22(2)：36-39.

[166] 罗家德. 社会网络分析讲义[M]. 北京：社会科学文献出版社. 2005. 132-184.

[167] 骆小平. “智慧城市”的内涵论析[J]. 城市管理与科技. 2010，(6)：34-37.

[168] 毛光烈．智慧城市需“标准化”建设[J]．信息化建设．2012，23(10)：10-12.

[169] 苗圩．建设现代信息技术产业体系[J]．求是．2012(23)：39-42.

[170] 倪冬梅，易兰丽，王理．国外智慧城市评估体系介绍(一)——“智慧社区”评估体系[EB/OL]．http：//www．cstc．org．cn/plus/view．php?aid＝5441，2012-08-13/2013-03-08.

[171] 欧阳俊，李康杓．中国汽车零部件购销网络的块段模型分析[J]．统计研究．2006(4)：32-38.

[172] 钱学森，于景元，戴汝为．一个科学新领域——开放的复杂巨系统及其方法论[J]．自然杂志．1990，13(1)：3-10.

[173] 乔晓春 ，黄衍华．中国跨省流动人口状况——基于六普数据的分析[J]．人口与发展．2013，19(1)：13-28.

[174] 任红娟，张志强．2010．基于文献内容和链接融合的知识结构划分方法研究进展[J]．情报理论与实践．2010，33(4)：124-128.

[175] 上海浦东智慧城市研究院．首个智慧城市指标体系亮相浦东[J]．中国公共安全(综合版)．2011，189(8)：120.



[176] 尚进．从“社会-科技”角度看智慧城市建设的实质与创新[J]．中国信息界．2012，(9)：66-69.

[177] 史美强，王光旭．台湾府际财政治理的竞合关系：一个网络分析的实证研究[J]．公共行政学报(台)．2008(28)：39-83.

[178] 宋刚，唐蔷，陈锐等．复杂性科学视野下的科技创新[J]．科学对社会的影响．2008(2)：28-33.

[179] 宋健，侯佳伟．流动人口管理：北京市相关政策法规的演变[J]．市场与人口分析．2007，13(3)：14-23.

[180] 宋丽华，姜家轩，张建成等．黄河三角洲云计算平台关键技术的研究[J]，计算机技术与发展．2011(6).

[181] 宋新平，吴晓伟，刘竞．基于信息融合和综合集成研讨厅混合的企业竞争情报系统[J]．图书情报工作．2009，53(22)：76-79.

[182] 孙吉贵，刘杰，赵连宇．聚类算法研究．软件学报．2008，19(1)：48-61.

[183] 孙茂源，于翔，魏以鹏．数字化城市物联网信息平台设计，物联网技术．2011(10).

[184] 唐东明. 聚类分析及其应用研究[D]. 成都：电子科技大学. 2010.

[185] 唐卫平，颜冰. 2005. 多传感器信息融合技术在网络雷阵中的应用[J]. 水雷战与舰船防护. 2005(2)：25-29.

[186] 万碧玉，姜栋，周微茹. 国家智慧城市试点与标准化建设探索[J]. 中兴通讯技术. 2014，20(4)：2-6.

[187] 王恩峰，郑尹茹. 线上与线下世界的交错：校园线上学习的社会网络分析[J]. 咨询社会研究(台). 2005(1)：155-192.

[188] 王洪斌，刘少岗，李瑶瑶等. 基于自适应模糊聚类的 T-S 模糊辨识方法[J]. 模糊系统与数学. 2014，28(5)：137-142.

[189] 王林，戴冠中. 复杂网络的 Scale-free 性、Scale-free 现象及其控制[M]. 北京：科学出版社. 2009.

[190] 王陆. 虚拟学习社区的社会网络分析[J]. 中国电化教育. 2009(2)：5-11.

[191] 王贤文，刘则渊，栾春娟，等. SSCI 数据库中的人文地理学期刊分析[J]. 地理学报. 2009(2)：243-252.

[192] 维克托·迈尔-舍恩伯格，肯尼思·库克耶著. 盛杨燕，周涛译. 大数据时代——生活、工作与思维的变革[M]. 杭州：浙江人民出版社. 2013.

[193] 翁顺裕. 从技术结构探讨技术群集之间的发展关系[J]. 科技发展政策导报. 2009(1)：3-17.

[194] 巫细波，杨再高. 智慧城市理念与未来城市发展[J]. 城市发展研究. 2010，17(11)：56-60.

[195] 吴余龙，艾浩军. 智慧城市：物联网背景下的现代城市建设之道[M]. 北京：电子工业出版社. 2011.

[196] 武传坤，物联网安全架构初探[J]. 战略与决策研究. 2010，25(4).

[197] 肖宇，于剑. 基于近邻传播算法的半监督聚类. 软件学报. 2008，19(11)：2803-2813.

[198] 杨再高. 智慧城市发展策略研究[J]. 科技管理研究. 2012，32(7)：20-24.

[199] 姚华松. 中国流动人口研究进展[J]. 城市问题. 2008，155(6)：69-76.

[200] 袁媛，高林，王潮阳等. 物联网与 SOA 在智慧城市的应用研究[J]. 信息技术与标准化. 2012(7).



[201] 张海涛,张永奎. 物联网体系架构与核心技术[J]. 长春工业大学学报(自然科学版).2012(4).

[202] 张建勋,古志民,郑超. 云计算研究进展综述[J]. 计算机研究进展综述. 2010(2).

[203] 张凌云,黎巎,刘敏. 智慧旅游的基本概念与理论体系[J]. 旅游学刊.2012(5).

[204] 张贤明,NASA 火星漫游车项目使用云计算[J]. 中国航天.2011(5).

[205] 张永刚,岳高峰. 我国智慧城市标准体系研究初探[J]. 标准科学.2013(11):14-18.

[206] 赵国栋,易欢欢,糜万军等. 大数据时代的历史机遇——产业变革与数据科学[M]. 北京:清华大学出版社.2013.

[207] 赵秋成,李怀. 目前我国人口流动的经济学意义[J]. 北方论丛.1996,137(3):68-73.

[208] 赵艳玲,李战宝. 云计算及其安全在美国的发展研究[J]. 信息网络安全.20U(10).

[209] 郑大永. 建设智慧城市要实现核心特征[A]. 四川省通信学会. 四川省通信学会2011 年学术年会论文集[C]. 四川省通信学会.2011:3.

[210] 周宏仁. 全面提高信息化水平[J]. 中国信息界.2011(11):9-16.

[211] 朱杰. 长江三角洲人口迁移空间格局、模式及启示[J]. 地理科学进展.2009,28(3):353-361.

[212] 朱子华. 图书信息融合系统的综合集成研讨厅机制研究[J]. 情报杂志.2007(8):88-89.

[213] 邹佳佳,马永俊. 智慧城市内涵与智慧城市建设[J]. 无线互联科技.2012,(4):69-70.